THE ANNALS OF MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

[illegible]

23, No. 4

THE [illegible] OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of [illegible] Statistics in March, June, September and December at Baltimore, Maryland.

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Department of Mathematics, Michigan State College, East Lansing, Michigan
K. J. Arnold, Secretary-T[illegible]
This address should be [illegible] communications concerning membership, subscriptions, [illegible], back numbers, etc., but not for [illegible]. Changes in mailing address which are to [illegible] for a given issue should be reported to the Secretary [illegible] before the [illegible] of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
After December 31, 195[illegible] correspondence should be directed to the Editor, [illegible], Statistical Laboratory, University of California, Berkeley 4, California. A manuscript should [illegible] double-spaced with wide margins, and the [illegible] (preferably with one additional [illegible]). Footnotes [illegible] reduced to an absolute minimum. [illegible] should be [illegible] drawn twice the size they are to be printed. Authors are [illegible] to keep in mind typographical difficulties of [illegible] mathematical formulae.
Authors will ordinarily receive [illegible] reprints without covers will be [illegible]. Additional reprints and covers furnished at cost.

Subscription Price: $[illegible] per year in the U.S. [illegible]. Single issues $[illegible]. Back numbers are [illegible] at the rate of $[illegible] per issue.

Composed and printed by [illegible]
[illegible], Inc., [illegible]

Entered as second-class matter at the Post Office [illegible]

A MEASURE OF ASYMPTOTIC EFFICIENCY FOR TESTS OF A HYPOTHESIS BASED ON THE SUM OF OBSERVATIONS[1]

By Herman Chernoff

University of Illinois and Stanford University

1. Summary. In many cases an optimum or computationally convenient test of a simple hypothesis H_0 against a simple alternative H_1 may be given in the following form. Reject H_0 if $S_n = \sum_{j=1}^{n} X_j \leqq k$, where $X_1, X_2, \cdots, X_n$ are n independent observations of a chance variable X whose distribution depends on the true hypothesis and where k is some appropriate number. In particular the likelihood ratio test for fixed sample size can be reduced to this form. It is shown that with each test of the above form there is associated an index ρ. If ρ_1 and ρ_2 are the indices corresponding to two alternative tests $e = \log \rho_1/\log \rho_2$ measures the relative efficiency of these tests in the following sense. For large samples, a sample of size n with the first test will give about the same probabilities of error as a sample of size en with the second test.

To obtain the above result, use is made of the fact that $P(S_n \leqq na)$ behaves roughly like m^n where m is the minimum value assumed by the moment generating function of $X - a$.

It is shown that if H_0 and H_1 specify probability distributions of X which are very close to each other, one may approximate ρ by assuming that X is normally distributed.

2. Introduction. The problem of the efficiency of a test is of relevance to statisticians who are faced with either of the following two problems. The first problem is that of the design of an experiment. The second problem is that of deciding which test combines computational feasibility and efficiency per observation. The measure of efficiency with which we shall deal is especially relevant to problems which involve large samples whose size is determined by the experimenter.

The motivation for the results of this paper may be seen by considering the following simple example. Suppose that under the hypothesis H_i,

$$P(X = 1) = p_i, \qquad P(X = 0) = 1 - p_i, \qquad i = 0, 1, \qquad p_1 > p_0. \tag{2.1}$$

Then the likelihood ratio test reduces to that of rejecting H_0 if $S_n = \sum_{j=1}^{n} X_j$ exceeds some number k. If $n = 400$, $p_0 = .4$, $p_1 = .5$, and $k = 180$, one may reliably proceed to compute the probabilities of error by using the normal approximation to the distribution of S_n. On the other hand, if n is very large, (say, 1,000,000) the difference between the means of S_n under H_0 and H_1 is so large

[1] This paper was prepared with the support of the Office of Naval Research.

compared to the standard deviation of S_n (ratio of 200) that the probabilities to be computed correspond to the extreme tails of the distributions of S_n and the normal approximation is inapplicable. We note that this objection would not be serious for $n = 1{,}000{,}000$ if p_0 were very close to p_1 (say, $p_0 = .499$) for then

$$(np_1 - np_0)/\sqrt{np_iq_i} = \sqrt{n}(p_1 - p_0)/\sqrt{p_iq_i} = 2.$$

This situation immediately gives rise to the question of what is the behaviour of the probability distribution of S_n in the tails of its distribution. This question was treated by H. Cramér [1] and is considered in Section 3 where Theorem 1 states that if $a \leqq E(X)$, $P(S_n \leqq na)$ is roughly like m^n where m is the minimum value assumed by the moment generating function of $X - a$. In Section 4 this result is applied to obtain a theorem which states the following result: If k is selected to minimize $\beta + \lambda\alpha$ where λ is some given positive number and $\alpha = P(S_n > k \mid H_0)$ and $\beta = P(S_n \leqq k \mid H_1)$ are the probabilities of error, the minimum value of $\beta + \lambda\alpha$ behaves roughly like ρ^n, where ρ does not depend on λ. Now the notion of efficiency is immediately suggested by the equation

$$\rho_1^{n_1} = \rho_2^{n_2}. \tag{2.2}$$

We may note that in the above example, one may be justified in using the normal approximation to the distribution of S_n for relatively large n if $p_1 - p_0$ is small. This tends to suggest that, if the hypotheses H_0 and H_1 are very "close" to each other, ρ may be approximated by assuming X to be normally distributed. This conjecture is in fact borne out by the theorems of Section 5.

3. The distribution of S_n in the tails. In this section we shall discuss the distribution in the tails of the sum of n independent observations on a chance variable X. Excellent results on this problem were obtained by H. Cramér [1] under the conditions that the moment generating function $M(t)$ of X exists (finite) for some interval $-A < t < A$, and that the cumulative distribution function of the chance variables have an absolutely continuous component. This latter condition is not satisfied by discrete distributions. This condition was imposed in order to apply a bound on the error of the normal approximation to the distribution of a sum of chance variables. C. G. Esseen [2] obtained this bound using only the (finite) existence of third order moments. For the case in which we are interested (i.e., $P(S_n \leqq na)$), the former condition may also be relaxed so that $M(t)$ exists (finite) for $-A < t \leqq 0$ if $a < E(X)$.

Since the results of Cramér are extremely more powerful that we require here and the (finite) existence of third order moments is not necessary for the results that we desire, we shall state and briefly outline a proof of Theorem 1. Before doing this we shall first formally state some notation and lemmas which we shall use throughout this paper. These lemmas state known results which are rather obvious, depending mainly on Lebesgue's Theorem on integration of monotone sequences [3].

NOTATION 1. *S_n is the sum of n independent observations $X_1, X_2, \cdots, X_n$ on a chance variable X with moment generating function $M(t) = E(e^{tX})$ and cumulative distribution function $F(x) = P(X \leqq x)$. Let*

$$m(a) = \inf E(e^{t(X-a)}) = \inf e^{-at}M(t) \tag{3.1}$$

(infimum with respect to real values of t).

Unless otherwise specified we shall say that an expectation exists if it is $+\infty$ or if it is $-\infty$. We shall say that $E(g(X))$ fails to exist if both

$$\int_{g(x)<0} g(x)\, dF(x) = -\infty \quad \text{and} \quad \int_{g(x)>0} g(x)\, dF(x) = +\infty.$$

We shall denote by $f(\infty)$ the limit of $f(x)$ as x approaches ∞.

LEMMA 1. *$M(t)$ attains its minimum value $m(0)$. This value is attained for finite t unless $P(X > 0) = 0$ or $P(X < 0) = 0$. In that event $m(0) = P(X = 0)$.*

LEMMA 2. *If $P(X \leqq 0) > 0$ and $P(X \geqq 0) > 0$, then $m(0) > 0$.*

LEMMA 3. *For all t in the interior of the interval of finite existence of $M(t)$*

$$\frac{dM}{dt} = \int_{-\infty}^{\infty} xe^{tx}\, dF(x) \tag{3.2}$$

and

$$\frac{d^2M}{dt^2} = \int_{-\infty}^{\infty} x^2e^{tx}\, dF(x) \geqq 0. \tag{3.3}$$

Furthermore, $\dfrac{d^2M}{dt^2} = 0$ if and only if $P(X = 0) = 1$.

LEMMA 4. *If $u_1(t), u_2(t), \cdots, u_n(t), \cdots$ is a nondecreasing sequence of functions continuous in the closed interval $[a, b]$,*

$$\lim_{n\to\infty} [\inf_{a\leqq t\leqq b} u_n(t)] = \inf_{a\leqq t\leqq b} [\lim_{n\to\infty} u_n(t)]. \tag{3.4}$$

This statement applies to the extended case where $u_n(t)$ may take on the value ∞ and to the case where $a = -\infty$ providing $u_n(-\infty) = \lim_{t\to-\infty} u_n(t)$.

THEOREM 1. *If $E(X) > -\infty$ and $a \leqq E(X)$, then*

$$P(S_n \leqq na) \leqq [m(a)]^n. \tag{3.5}$$

If $E(X) < +\infty$ and $a \geqq E(X)$, then

$$P(S_n \geqq na) \leqq [m(a)]^n. \tag{3.6}$$

If $0 < \epsilon < m(a)$ ($E(X)$ need not exist),

$$\lim_{n\to\infty} \frac{(m(a) - \epsilon)^n}{P(S_n \leqq na)} = 0 \tag{3.7}$$

and

$$\lim_{n\to\infty} \frac{(m(a) - \epsilon)^n}{P(S_n \geqq na)} = 0. \tag{3.8}$$

PROOF. We present here a brief sketch of a proof of Theorem 1. We note first that it suffices to prove (3.5) for $a = 0 \leqq E(X)$, and (3.7) for $a = 0$. Using the extended Tchebycheff inequality [4]

$$E(e^{tS_n}) = [M(t)]^n \geqq P[S_n \leqq 0] \qquad \text{for } t \leqq 0. \tag{3.9}$$

Hence

$$P[S_n \leqq 0] \leqq [\inf_{t \leqq 0} M(t)]^n. \tag{3.10}$$

But $a \leqq E(X)$ implies that

$$\inf_{t \leqq 0} M(t) = \inf M(t) = m(0). \tag{3.11}$$

To establish equation (3.7) we note that it is sufficient to treat the case $a = 0$. Then we see that the cases where $P(X > 0) = 0$ and where $P(X < 0) = 0$ are trivial. Hereafter we shall assume that $P(X > 0) > 0$ and $P(X < 0) > 0$.

We shall now treat the discrete (but not necessarily bounded) case where $P(X = x_i) = p_i > 0$, $i = 1, 2, \cdots$. Given $\epsilon > 0$, one may select an integer r so that

$$\min (x_1, x_2, \cdots, x_r) < 0 < \max (x_1, x_2, \cdots, x_r) \tag{3.12}$$

and

$$\inf \left\{ \sum_{i=1}^{r} e^{tx_i} p_i \right\} > \inf \left\{ \sum_{i=1}^{\infty} e^{tx_i} p_i \right\} - \frac{\epsilon}{2}. \tag{3.13}$$

In fact, let

$$m^* = \sum_{i=1}^{r} e^{t^* x_i} p_i = \inf \left\{ \sum_{i=1}^{r} e^{tx_i} p_i \right\}. \tag{3.14}$$

For this discrete case it now suffices to show that for sufficiently large n there are r positive integers $n_1, n_2, \cdots, n_r$ such that

$$\sum_{i=1}^{r} n_i = n, \tag{3.15}$$

$$\sum_{i=1}^{r} n_i x_i \leqq 0, \tag{3.16}$$

$$P(n_1, n_2, \cdots, n_r) = \frac{n!\, p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}}{n_1!\, n_2! \cdots n_i!} > \left(m^* - \frac{\epsilon}{2}\right)^n. \tag{3.17}$$

For large $n_1, n_2, \cdots, n_r$ (not necessarily integers) Stirling's Formula gives us

$$P(n_1, n_2, \cdots, n_r) \geqq \left\{\prod_{i=1}^{r}\left(\frac{np_i}{n_i}\right)^{n_i}\right\} \cdot \frac{1}{n^{r/2}}. \tag{3.18}$$

Now

$$Q(n_1, n_2, \cdots, n_r) = \prod_{i=1}^{r}\left(\frac{np_i}{n_i}\right)^{n_i} \tag{3.19}$$

can be shown by the method of Lagrange multipliers to attain a maximum of $(m^*)^n$ subject to the restrictions

$$\sum_{i=1}^{r} n_i = n, \tag{3.20}$$

$$\sum_{i=1}^{r} n_i x_i = 0, \tag{3.21}$$

$$n_i > 0, \qquad i = 1, 2, \cdots, r, \tag{3.22}$$

and the maximizing values of $n_1, n_2, \cdots, n_r$ are

$$n_i^{(0)} = np_i e^{t^* x_i}/m^*. \tag{3.23}$$

Assuming that $x_1 \leqq x_i$ for $i \leqq r$, we let

$$n_i^{(1)} = [n_i^{(0)}], \qquad 2 \leqq i \leqq r, \tag{3.24}$$

$$n_1^{(1)} = n - \sum_{i=2}^{r} n_i^{(1)}, \tag{3.25}$$

where $[n_i^{(0)}]$ represents the greatest integer less than or equal to $n_i^{(0)}$. Then for large n, the $n_i^{(1)}$ are large positive integers adding up to n for which

$$\textstyle\sum_{i=1}^{r} n_i^1 x_i \leqq 0$$

and

$$Q(n_1^{(1)} n_2^{(1)}, \cdots, n_r^{(1)}) \geqq \left(\frac{p_1}{n}\right)^r (m^*)^n \tag{3.26}$$

and

$$P(n_1^{(1)}, n_2^{(2)}, \cdots, n_r^{(1)}) \geqq \frac{(m^*)^n p_1^r}{n^{3r/2}} \geqq \left(m^* - \frac{\epsilon}{2}\right)^n, \tag{3.27}$$

which was to be shown.

We shall now treat the general case. Let

$$X^{(s)} = \frac{i}{s} \quad \text{if} \quad \frac{i-1}{s} < X \leqq \frac{i}{s}, \qquad i = \cdots, -1, 0, 1, \cdots, \tag{3.28}$$
$$s = 1, 2, \cdots.$$

If $S_n^{(s)}$ represents the sum of the $X^{(s)}$ for n independent observations

$$P(S_n \leqq 0) \geqq P(S_n^{(s)} \leqq 0) \tag{3.29}$$

and

$$M^{(s)}(t) = E(e^{tX^{(s)}}) \geqq e^{-|t|/s}M(t). \tag{3.30}$$

Since $P(X > 0) > 0$ and $P(X < 0) > 0$, $M(t)$ attains its minimum for a finite value of t and hence there is an s sufficiently large so that

$$\inf \{M^{(s)}(t)\} \geqq \inf \{M(t)\} - \tfrac{1}{2}\epsilon. \tag{3.31}$$

Our theorem follows from the result for the discrete case and equation (3.29).

4. The measure of asymptotic efficiency. In this section some elementary monotonicity and continuity properties of $m(a)$ are obtained. These properties are then used to obtain an index ρ for a test. This index has the property that if k is chosen to minimize

$$\beta + \lambda\alpha = P[S_n \leqq k \mid H_1] + \lambda P[S_n > k \mid H_0], \tag{4.1}$$

the minimum value of $\beta + \lambda\alpha$ is roughly about ρ^n. Furthermore, ρ is independent of λ. From this it is easily seen that if ρ_1 and ρ_2 are the indices of two tests, $\log \rho_1/\log \rho_2$ is an appropriate measure of the relative efficiency of these tests.

NOTATION 2. *Let a_e be defined by*

$$P(X < a_e) = 0 \tag{4.2}$$

and

$$P(X < a_e + \epsilon) > 0 \qquad \textit{for every } \epsilon > 0. \tag{4.3}$$

Let $t(a)$ be given by

$$m(a) = e^{-at(a)}M[t(a)]. \tag{4.4}$$

Note that Lemma 1 implies that $t(a)$ exists and that Lemma 3 implies that $t(a)$ is unique unless $P(X = a) = 1$.

LEMMA 5. *If $E(X) > -\infty$ and $M(t) = \infty$ for $t < 0$, then $t(a) = 0$ and $m(a) = 1$ for $a \leqq E(X)$.*

PROOF. From the proof of (3.5), it follows that $t(a) \leqq 0$ for $a \leqq E(X)$. Lemma 5 follows immediately.

LEMMA 6. *If $M(t) < \infty$ for some $t < 0$, then $E(x) > -\infty$. Furthermore,*

$$m(a) = 0, \qquad a < a_e, \tag{4.5}$$

$$m(a_e) = P(X = a_e), \tag{4.6}$$

and

$$m[E(X)] = 1. \tag{4.7}$$

Also, $m(a)$ is continuous and strictly monotone increasing for $a_e \leqq a \leqq E(X)$.

PROOF. That $E(X)$ exists (finite) or is $+\infty$ is apparent. For $a < a_e$ and $t < 0$,

$$\int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) < e^{t(a_e-a)}. \tag{4.8}$$

Hence $m(a) = 0$. Now we note that if a_e is finite

$$P(X = a) \leqq \int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x), \tag{4.9}$$

$$\lim_{t\to-\infty} \int_{-\infty}^{\infty} e^{t(x-a_e)}\, dF(x) = P(X = a_e) \tag{4.10}$$

and hence $m(a_e) = P(X = a_e)$. If $a_e = -\infty$,

$$\lim_{a\to-\infty} \int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) = 0, \qquad t < 0, \tag{4.11}$$

so that $\lim_{a\to-\infty} m(a) = 0$. Now we note that

$$\lim_{t\to0-} \frac{d}{dt}\,[e^{-at} M(t)] = \int_{-\infty}^{\infty} (x - a)\, dF(x). \tag{4.12}$$

Since $(d^2/dt^2)[e^{-at}M(t)] > 0$, unless $P(X = a) = 1$ in which case Lemma 6 is valid, it follows that $t(a) < 0$ for $a < E(X)$ and $t(a) = 0$ for $a = E(X)$. Hence $m[E(X)] = 1$, and $m(a) < 1$ for $a < E(X)$.

We shall now show that for $a_e < a < E(X)$, $t(a)$ is finite and a non-decreasing function of a, while $m(a)$ is strictly increasing for $a_e \leqq a < E(X)$. The finiteness of $t(a)$ follows from

$$\int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x) \geqq P(X < a - \epsilon)e^{-\epsilon t} \tag{4.13}$$

for $t < 0$, $\epsilon > 0$. Therefore,

$$m(a - h) \leqq \int_{-\infty}^{\infty} e^{t(a)(x-a+h)}\, dF(x) < m(a), \qquad h > 0, \tag{4.14}$$

and furthermore

$$\int_{-\infty}^{\infty} e^{t'(x-a+h)}\, dF(x) \geqq \int_{-\infty}^{\infty} e^{t(a)(x-a+h)}\, dF(x) \tag{4.15}$$

for $t' > t(a)$, $h > 0$.

It suffices now to show that $m(a)$ is continuous on the right for $a < E(X)$ and continuous on the left for $a_e < a \leqq E(X)$. Given $a < E(X)$ and $\epsilon > 0$, there is a finite t' so that

$$\int_{-\infty}^{\infty} e^{t'(x-a)}\, dF(x) \leqq m(a) + \epsilon, \tag{4.16}$$

$$\lim_{h\to 0+} m(a+h) \leqq \lim_{h\to 0+} \int_{-\infty}^{\infty} e^{t'(x-a-h)}\, dF(x) \tag{4.17}$$
$$\leqq m(a) + \epsilon.$$

Given a so that $a_e < a \leqq E(X)$, there is an $h_1 > 0$ so that $a_e < a - h_1$. For $t(a - h_1) \leqq t \leqq t(a)$, $\int_{-\infty}^{\infty} e^{t(x-a+h)}\, dF(x)$ converges uniformly to $\int_{-\infty}^{\infty} e^{t(x-a)}\, dF(x)$ as $h \to 0+$. Hence $\lim_{h\to 0+} m(a-h) \geqq m(a)$.

NOTATION 3. *H_0 and H_1 are two hypotheses which specify the distribution of X so that $\mu_0 = E(X \mid H_0) \leqq \mu_1 = E(X \mid H_1)$. For each value of a we consider a test which consists of rejecting H_0 if $S_n > na$. Let $\alpha = P(S_n > na \mid H_0)$, $\beta = P(S_n \leqq na \mid H_1)$ and λ be any (finite) positive number given in advance. Let*

$$\rho(a) = \max\,[m_0(a), m_1(a)], \tag{4.18}$$

where

$$m_i(a) = \inf E(e^{t(X-a)} \mid H_i), \qquad i = 0, 1. \tag{4.19}$$

Furthermore, let us define the index of the test determined by X by

$$\rho = \inf_{\mu_0 \leqq a \leqq \mu_1} \rho(a). \tag{4.20}$$

We note that in the event that it is desired to use a test where we reject H_0 if $S_n \geqq na$, one may replace X by $-X$ and interchange H_0 and H_1. The value of ρ is not affected by this transformation.

The customary procedure of minimizing β for a fixed value of α does not seem very appropriate when the sample size approaches infinity. We shall instead deal with test which minimize $\beta + \lambda\alpha$ for some fixed value of λ, $0 < \lambda < \infty$. Such a test is a "Bayes Solution" corresponding to some a priori probability of H_0 which depends on λ. The study of Bayes Solutions may here be justified on grounds not involving any belief in a priori probabilities. In particular, if it is desired to minimize some function $F(\alpha, \beta)$ for large samples and neither $\partial F/\partial\alpha$ nor $\partial F/\partial\beta$ vanish at $\alpha = \beta = 0$, the minimizing test will correspond to a Bayes Solution where λ is close to $\dfrac{\partial F(0,0)}{\partial\beta} \Big/ \dfrac{\partial F(0,0)}{\partial\alpha}$.

THEOREM 2. *Given ϵ and λ, $\epsilon > 0$ and $0 < \lambda < \infty$, then*

$$\lim_{n\to\infty} \left\{ \inf_a\, (\beta + \lambda\alpha)/(\rho + \epsilon)^n \right\} = 0 \tag{4.21}$$

and if $0 < \epsilon < \rho$

$$\lim_{n\to\infty} \left\{ \inf_a\, (\beta + \lambda\alpha)/(\rho - \epsilon)^n \right\} = \infty. \tag{4.22}$$

PROOF. There is a value a_0 of a so that $\rho(a_0) \leqq \rho + \epsilon/2$. Applying Theorem 1, equation (4.21) follows immediately. Now we note that

$$\beta = P(S_n \leqq na \mid H_1) \geqq P(S_n \leqq na_1 \mid H_1), \qquad a \geqq a_1, \tag{4.23}$$

$$\inf_{a \geqq a_1} (\beta + \lambda\alpha) \geqq P(S_n \leqq na_1 \mid H_1) \tag{4.24}$$

$$\alpha = P(S_n > na \mid H_0) \geqq P(S_n \geqq na_1 \mid H_0), \qquad a < a_1, \tag{4.25}$$

$$\inf_{a < a_1} (\beta + \lambda\alpha) \geqq \lambda P(S_n \geqq na_1 \mid H_0). \tag{4.26}$$

Theorem 1 gives us our result as soon as we show the existence of an a_2 in $[\mu_0, \mu_1]$ so that both $m_0(a_2) \geqq \rho$ and $m_1(a_2) \geqq \rho$. To this end we consider

$$F = \{a : m_1(a) \geqq \rho, \mu_0 \leqq a \leqq \mu_1\}. \tag{4.27}$$

The set F is not empty because $m_1(\mu_1) = 1 \geqq \rho$. Let $a_2 = g.\ l.\ b.\ F$. By continuity on the right $m_1(a_2) \geqq \rho$. Also $m_1(a) < \rho$ for $a < a_2$. Hence $m_0(a) \geqq \rho$ if $\mu_0 \leqq a < a_2$. Since $m_0(a)$ is continuous on the left for $a > \mu_0$, $m_0(a_2) \geqq \rho$ if $a_2 > \mu_0$. Furthermore, if $a_2 = \mu_0$, $m_0(a_2) = 1 \geqq \rho$.

NOTATION 4. *Let ρ_1 and ρ_2 represent the indices of two tests T_1 and T_2, respectively. We define the asymptotic relative efficiency of T_1 to T_2 by*

$$e = \log \rho_1 / \log \rho_2, \tag{4.28}$$

where e is undefined if $\rho_1 = \rho_2 = 1$. For test T_i, n_i is the sample size and

$$\gamma_i = \inf (\beta + \lambda\alpha) \tag{4.29}$$

is a function of n_i and λ.

The appropriateness of the use of e as a measure of efficiency derives from the following theorem, which is an immediate consequence of Theorem 2.

THEOREM 3. *If* $\lim_{n_1, n_2 \to \infty} \frac{n_2}{n_1} < e(>e)$, *then* $\lim_{n_1, n_2 \to \infty} \frac{\gamma_1}{\gamma_2} = \infty\,(=0)$.

Note that e does not depend on λ.

5. Some examples. In this section we shall determine the behaviour of $m(a)$ and ρ for a few simple examples.

EXAMPLE 1. Let X be normally distributed with mean μ_i and variance σ_i^2 under hypothesis H_i, $i = 0, 1 (\mu_0 < \mu_1)$. Then

$$e^{-at} M_i(t) = e^{(\mu_i - a)t + \frac{1}{2}\sigma_i^2 t^2} \tag{5.1}$$

$$m_i(a) = e^{\frac{1}{2}(\mu_i - a)^2/\sigma_i^2}, \tag{5.2}$$

$$\rho = \rho\left(\frac{\sigma_1\mu_0 + \sigma_0\mu_1}{\sigma_1 + \sigma_0}\right) = e^{-\frac{1}{2}((\mu_1 - \mu_0)/(\sigma_1 + \sigma_0))^2}. \tag{5.3}$$

Of course this index applies to a test which is not the likelihood ratio test unless $\sigma_1 = \sigma_0$. The computational problem in obtaining the index of the likelihood ratio test is considerable. However, the results in Section 6 may be easily applied to the likelihood ratio test if $\mu_1 - \mu_0$ and $\sigma_1 - \sigma_0$ are small.

EXAMPLE 2. Let X/σ_i^2 have a chi-square distribution with r degrees of freedom under hypothesis H_i, $i = 0, 1(\sigma_0^2 < \sigma_1^2)$. Then

$$e^{-at}M_i(t) = e^{-at}(1 - 2\sigma_i^2 t)^{-\frac{1}{2}r} \tag{5.4}$$

$$\log m_i(a) = -\tfrac{1}{2}\left[\frac{a}{\sigma_i^2} + r \log \frac{r\sigma_i^2}{a} - r\right], \tag{5.5}$$

$$\log \rho = -\tfrac{1}{2}r[\delta - 1 - \log \delta], \tag{5.6}$$

where

$$\delta = (\log \tau)/(\tau - 1), \tag{5.7}$$

$$\tau = \sigma_0^2/\sigma_1^2 . \tag{5.8}$$

Note that as τ approaches 1, $\log \rho \approx -r(\tau - 1)^2/16$.

EXAMPLE 3. Let X have the binomial distribution so that

$$P(X = j \mid H_i) = \binom{r}{j} p_i^j q_i^{r-j}, \tag{5.9}$$
$$q_i = 1 - p_i, \quad i = 0, 1, \quad j = 0, 1, \cdots, r(p_0 < p_1).$$

Then

$$e^{-at}M_i(t) = e^{-at}(p_i e^t + q_i)^r, \tag{5.10}$$

$$\log m_i(a) = (r - a) \log [rq_0/(r - a)] + a \log [rp_0/a], \tag{5.11}$$

$$\log \rho = r\{(1 - c) \log [q_0/(1 - c)] + c \log [p_0/c]\}, \tag{5.12}$$

where

$$c = \frac{\log (q_0/q_1)}{\log (q_0/q_1) + \log (p_1/p_0)}, \; p_0 < c < p_1 . \tag{5.13}$$

Note that as p_1 approaches p_0, $\log \rho \approx -r(p_1 - p_0)^2/8p_0q_0$.

6. Normality approximation. In this section we shall develop some results concerning the conjecture made in the introduction, that if the hypotheses H_0 and H_1 are very close to one another one may approximate ρ by assuming that X is normally distributed. To this end we shall first investigate more closely the behaviour of $m(a)$ and $t(a)$.

NOTATION 5. *Let $N(t) = E(Xe^{tX})$ and $P(t) = E(X^2e^{tX})$. Let*
$$t_0 = \text{glb}\{t : M(t) < \infty\}$$

and if $t_0 < 0$ let

$$a_0 = \inf_{t_0 < t < 0} N(t)/M(t). \tag{6.1}$$

Note that if $E(X) > -\infty$, $a_e < E(X)$ except in the case where $P(X = a_e) = 1$. Furthermore, if $a_e > -\infty$, then $t_0 = -\infty$.

LEMMA 7. *If $M(t) < \infty$ for some $t < 0$, and $a_0 < E(X)$, then for $a_0 < a < E(X)$*

$$a = N(t(a))/M(t(a)), \tag{6.2}$$

$$\frac{d[\log m(a)]}{da} = -t(a), \tag{6.3}$$

$$\frac{dt(a)}{da} = \frac{M(t)^2}{M(t)P(t) - N(t)^2}\bigg|_{t=t(a)} > 0. \tag{6.4}$$

If in addition $\mu = E(X)$ and $\sigma^2 = E[(X - \mu)^2]$ are finite, then (6.2), (6.3), and (6.4) *hold for $a_0 < a \leqq E(X)$, giving*

$$\frac{d}{da}[\log m(a)]\bigg|_{a=\mu} = 0 \tag{6.5}$$

and

$$\frac{dt(a)}{da}\bigg|_{a=\mu} = 1/\sigma^2. \tag{6.6}$$

PROOF. Suppose that $t_0 < t < 0$. Using Lemma 3, there is a unique a so that $t = t(a)$ and this value of a is obtained by

$$\frac{d}{dt}[e^{-at}M(t)] = \int_{-\infty}^{\infty}(x - a)e^{t(x-a)}\,dF(x) = 0 \tag{6.7}$$

and is given by

$$a = N(t)/M(t). \tag{6.8}$$

Considering a as a function of t we may differentiate

$$\frac{da}{dt} = \frac{M(t)P(t) - N(t)^2}{M(t)^2}. \tag{6.9}$$

Applying Schwarz' Inequality, the numerator is at least zero. It can vanish only if $Xe^{tX/2}$ and $e^{tX/2}$ are proportional with probability one. This can occur only if $P[X = a_0] = 1$. This case is excluded by the hypothesis $a_0 < E(X)$. Hence $(da/dt) > 0$. Furthermore, as $t \to 0$, $a \to E(X)$. Hence as t varies over $(t_0, 0)$ a ranges continuously (and monotonically) over $(a_0, E(X))$. Equations (6.2) and (6.4) are immediately valid. Equation (6.3) is obtained by differentiating with respect to a, $m(a) = E(e^{t(a)(X-a)})$. Equations (6.5) and (6.6) follow from the Lebesque Convergence Theorem [3] and the fact that if $f(x)$ is continuous at $x = a$, and $f'(x) \to b$ as $x \to a$, then $f'(a) = b$.

If ν_0 and ν_1 are any two probability measures defined on the same Borel Field, we may introduce the measure $\nu = (\nu_0 + \nu_1)/2$. A consequence of the Radon-Nikodym Theorem [3] is the existence of two densities f_0 and f_1 (unique except possibly on a set of ν measure zero) so that

$$\nu_i(E) = \int_E f_i(x)\,d\nu(x) \qquad i = 0, 1. \tag{6.10}$$

Hence, except on a set of ν measure zero, at least one of $f_0(x)$ and $f_1(x)$ are non-zero, and the log of the likelihood may be defined by $\log f_1(x) - \log f_0(x)$.

NOTATION 6. *The outcome of an experiment is denoted by Y and has a probability distribution given by equation (6.10) under hypothesis H_i. When an integration sign is unaccompanied by a region of integration it is to be understood that the region is the set of all possible values of Y. We shall deal with a chance variable X which is a function of Y. In particular the log of the likelihood ratio is defined by $\log f_1(Y) - \log f_0(Y)$.*

$$M_i(t) = E(e^{tX} \mid H_i), \tag{6.11}$$

$$N_i(t) = E(Xe^{tX} \mid H_i), \tag{6.12}$$

$$P_i(t) = E(X^2e^{tX} \mid H_i). \tag{6.13}$$

We use $m_i(a)$ and $t_i(a)$ to represent the functions $m(a)$ and $t(a)$ under hypothesis H_i.

LEMMA 8. *If X is the log of the likelihood ratio, $X \not\equiv 0$, and X is finite with probability one, then*

$$M_1(t) = M_0(t+1), \qquad N_1(t) = N_0(t+1). \tag{6.14}$$

As a varies from μ_0 to μ_1, $t_0(a)$ varies continuously from 0 to 1 and

$$t_0(a) = t_1(a) + 1, \tag{6.15}$$

$$\rho = \inf_{0<t<1} M_0(t). \tag{6.16}$$

PROOF. We note that

$$M_1(t) = \int \left[\frac{f_1(x)}{f_0(x)}\right]^t \frac{f_1(x)}{f_0(x)} f_0(x)\, d\nu(x) = M_0(t+1) \tag{6.17}$$

and

$$N_1(t) = \int \log\left[\frac{f_1(x)}{f_0(x)}\right]\left[\frac{f_1(x)}{f_0(x)}\right]^t \frac{f_1(x)}{f_0(x)} f_0(x)\, d\nu(x) = N_0(t+1). \tag{6.18}$$

It is evident that

$$M_0(0) = M_0(1) = M_1(0) = M_1(-1) = 1. \tag{6.19}$$

It follows that $N_0(t)$ is finite for $0 < t < 1$, that $\mu_1 > 0$, $\mu_0 < 0$, and

$$\lim_{t\to 0-} N_1(t) = N_1(0) = \mu_1, \tag{6.20}$$

$$\lim_{t\to -1+} N_1(t) = N_1(-1) = \mu_0. \tag{6.21}$$

Applying Lemma 7, we find that as a varies from μ_0 to μ_1, $t_1(a)$ varies continuously and (strictly) monotonically from -1 to 0. Similarly, $t_0(a)$ varies from 0 to 1. Applying equations (6.2), (6.17), and (6.18)

$$\frac{N_0[t_1(a)+1]}{M_0[t_1(a)+1]} = \frac{N_0[t_0(a)]}{M_0[t_0(a)]} = a \text{ for } \mu_0 < a < \mu_1. \tag{6.22}$$

Equation (6.15) follows.

Since $e^{-at_0(a)}M_0(t_0(a))$ and $e^{-at_1(a)}M_1(t_1(a))$ are both equal and continuous at $a = 0$, the monotonicity properties of Lemma 5 show that

$$\rho = M_0(t_0(0)) = m_0(0) = m_1(0) = \inf_{0<t<1} M_0(t), \tag{6.23}$$

$$\rho = \inf_{0<t<1} \int [f_1(x)]^t [f_0(x)]^{1-t}\, d\nu(x). \tag{6.24}$$

We are interested in likelihood ratio tests for which $\mu_0^2 + \sigma_0^2$ is very small. The following theorem applies to certain classes of tests. In this theorem we are interested in classes of tests where the log of the likelihood ratio has finite means. Hence the restriction of Lemma 8 that X is finite with probability one is automatically satisfied. However the case where X may assume the values $+\infty$ or $-\infty$ with positive probability is of some interest. For this case the above sort of reasoning applies except that all integrals must be taken over the set,

$$G = \{x: -\infty < \log f_1(x) - \log f_0(x) < \infty\}.$$

After the necessary modifications are made, it is seen that (6.24) is valid in general.

THEOREM 4. *If, for a class C of likelihood ratio tests,*

$$\begin{aligned} M_0(t) &= 1 + \mu_0 t + (\mu_0^2 + \sigma_0^2)t^2/2 + o(\mu_0^2 + \sigma_0^2), \qquad 0 < t < 1, \\ M_1(t) &= 1 + \mu_1 t + (\mu_1^2 + \sigma_1^2)t^2/2 + o(\mu_0^2 + \sigma_0^2), \qquad -1 < t < 0, \end{aligned} \tag{6.25}$$

then

$$\begin{aligned} \mu_1 &= \sigma_0^2/2 + o(\sigma_0^2), \\ \mu_0 &= -\sigma_0^2/2 + o(\sigma_0^2), \\ \sigma_1^2 &= \sigma_0^2 + o(\sigma_0^2), \end{aligned} \tag{6.26}$$

and

$$\begin{aligned} \rho &= e^{-\sigma_0^2/8} + o(\sigma_0^2), \\ \rho &= e^{\mu_0/4} + o(\mu_0), \\ \rho &= e^{-\frac{1}{2}[(\mu_1-\mu_0)/(\sigma_1+\sigma_0)]^2} + o\left[\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right]^2, \end{aligned} \tag{6.27}$$

PROOF. *Part 1.*

$$M_1(t - 1) = M_0(t), \qquad 0 < t < 1, \tag{6.28}$$

$$\left[\mu_1 - \frac{\mu_1^2 + \sigma_1^2}{2}\right] + t[(\mu_0 - \mu_1) + (\mu_1^2 + \sigma_1^2)] + \frac{t^2}{2}[\mu_0^2 + \sigma_0^2 - \mu_1^2 - \sigma_1^2] = o(\mu_0^2 + \sigma_0^2). \tag{6.29}$$

Hence

$$
\begin{aligned}
\mu_1^2 + \sigma_1^2 &= \mu_0^2 + \sigma_0^2 + o(\mu_0^2 + \sigma_0^2),\\
\mu_1 &= \sigma_1^2/2 + o(\mu_0^2 + \sigma_0^2),\\
\mu_0 &= \sigma_1^2/2 + o(\mu_0^2 + \sigma_0^2),\\
\sigma_1^2 &= \sigma_0^2 + o(\mu_0^2 + \sigma_0^2).
\end{aligned} \tag{6.30}
$$

Equations 6.26 follow immediately.

Part 2. By Lemma 8, $\rho = \inf_{0<t<1} M_0(t)$. Minimizing the quadratic approximation we obtain

$$
\rho = 1 - \frac{\mu_0^2}{2(\mu_0^2 + \sigma_0^2)} + o(\mu_0^2 + \sigma_0^2). \tag{6.31}
$$

Applying the results of Part 1, Part 2 follows immediately.

We may also be interested in tests of the form

$$
S_n = \sum_{j=1}^{n} X_j \leq k,
$$

where X is a less efficient statistic than the log of the likelihood ratio. Here again, given a class of tests, we may investigate the behaviour of ρ as the hypotheses get "close" together. For some such classes we state the following theorem.

THEOREM 5. *If, for a class C^* of tests,*

$$
\sigma_i^2 \frac{dt_i(a)}{da} = 1 + o(1) \tag{6.31}
$$

as $\mu_1 - \mu_0 \to 0$ for $\mu_0 < a < \mu_1$, $i = 0, 1$, then

$$
\rho = e^{-\frac{1}{2}[(\mu_1-\mu_0)/(\sigma_1+\sigma_0)]^2} + o\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right)^2. \tag{6.32}
$$

PROOF.

$$
\begin{aligned}
\log m_0(a) &= \frac{-(a - \mu_0)^2}{2}\left[\frac{1}{\sigma_0^2} + o\left(\frac{1}{\sigma_0^2}\right)\right],\\
\log m_1(a) &= \frac{-(a - \mu_1)^2}{2}\left[\frac{1}{\sigma_1^2} + o\left(\frac{1}{\sigma_1^2}\right)\right].
\end{aligned} \tag{6.33}
$$

Equating the main terms, one obtains

$$
\log \rho = -\frac{1}{2}\left(\frac{\mu_1 - \mu_0}{\sigma_1 + \sigma_0}\right)^2 + o\left(\frac{\mu_1 - \mu_0}{\sigma_1 - \sigma_0}\right)^2. \tag{6.34}
$$

It may be seen that the corresponding value of a satisfies $a \approx (\sigma_1\mu_0 + \sigma_0\mu_1)/(\sigma_1 + \sigma_0)$. Finally we note that equation (6.4) is useful in checking the applicability of Theorem 5.

7. Measures of information and divergence. In Section 4 the measure of efficiency e, was defined so that n observations for one test is equivalent to en observations for the second test (equivalent from the point of view of the criterion we used). It is evident that it would have been appropriate to use the following equation

$$I(X) = -\log \rho \tag{7.1}$$

to indicate that $-\log \rho$ may be used as a measure of the information per observation for a test based on sums of observations on X. (Here X denotes the two specified chance variables associated with H_0 and H_1, respectively.) In addition we may have written

$$D(Y) = -\log\left[\inf_{0<t<1}\int [f_1(x)]^t[f_0(x)]^{1-t}\,d\nu(x)\right] \tag{7.2}$$

to indicate that $-\log \rho$ for the likelihood ratio test may be used as a measure of the divergence between the two distributions associated with Y. Let (Y_1, Y_2) represent an observation consisting of independent observations on Y_1 and Y_2 respectively. Then it is easy to see from equation (6.24) that

$$D(Y_1, Y_2) \leqq D(Y_1) + D(Y_2) \tag{7.3}$$

and

$$D(Y, Y) = 2D(Y) \tag{7.4}$$

A measure of divergence used by Kullback and Leibler [5, 6], yields equality in the relation (7.3). The measure (7.2) and that used by Kullback and Leibler are basically two different functionals on the curve relating the type 1 and type 2 errors for likelihood ratio tests.

REFERENCES

[1] H. Cramér, "Sus un nouveau théorème-limite de la théorie des probabilités," *Actualités Scientifiques et Industrielles*, No. 736, Paris 1938.

[2] C. G. Esseen, "Fourier analysis of distribution functions, *Acta Mathematica*, Vol. 77, (1945) pp. 1–125.

[3] S. Saks, *Theory of the Integral*, 2nd ed., Warsaw, 1937.

[4] A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Julius Springer, Berlin, 1933.

[5] S. Kullback, and R. A. Leibler, "On information and sufficiency", *Annals of Math. Stat.*, Vol. 22, (1951) pp. 79–86,

[6] H. Jeffreys, *Theory of Probability*, 2nd ed., Oxford University Press, 1948.

ORTHOGONAL ARRAYS OF STRENGTH TWO AND THREE

BY R. C. BOSE AND K. A. BUSH

University of North Carolina and University of Illinois

1. Summary. Orthogonal arrays can be regarded as natural generalizations of orthogonal Latin squares, and are useful in various problems of experimental design. In this paper the known upper bounds for the maximum possible number of constraints for arrays of strength 2 and 3 have been improved, and certain methods for constructing these arrays have been given.

2. Introduction. A $k \times N$ matrix A, with entries from a set Σ of $s \geqq 2$ elements, is called an orthogonal array of strength t, size N, k constraints and s levels if each $t \times N$ submatrix of A contains all possible $t \times 1$ column vectors with the same frequency λ. The array may be denoted by (N, k, s, t). The number λ may be called the index of the array. Clearly $N = \lambda s^t$.

The set Σ will for convenience be taken as the set of integers $0, 1, 2, \cdots, s - 1$. For example the orthogonal array (18, 7, 3, 2) with index 2 is given below. It is easy to verify that in any 2×18 submatrix, each of the column vectors (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2) occurs twice.

$$\begin{matrix}
0&1&2&0&1&2&0&1&2&0&1&2&0&1&2&0&1&2\\
0&1&2&0&1&2&1&2&0&2&0&1&1&2&0&2&0&1\\
0&1&2&1&2&0&0&1&2&2&0&1&2&0&1&1&2&0\\
0&1&2&2&0&1&2&0&1&0&1&2&1&2&0&1&2&0\\
0&1&2&1&2&0&2&0&1&1&2&0&0&1&2&2&0&1\\
0&1&2&2&0&1&1&2&0&1&2&0&2&0&1&0&1&2\\
0&0&0&0&0&0&1&1&1&1&1&1&2&2&2&2&2&2
\end{matrix} \tag{2.0}$$

If the orthogonal array A is of strength t so is any subarray of k' rows (constraints) if $k' \leqq k$. Hence the non-existence of $(\lambda s^t, k', s, t)$ automatically implies the non-existence of $(\lambda s^t, k, s, t)$ if $k > k'$. Again if A is of strength t, it is also of strength t' for all $t' \leqq t$.

The optimum multifactorial designs considered by Plackett and Burman [1] are essentially orthogonal arrays of strength 2. They have shown that the maximum number of constraints k for an orthogonal array of size λs^2, s levels and strength 2, satisfies the inequality

$$k \leqq \left[\frac{\lambda s^2 - 1}{s - 1}\right], \tag{2.1}$$

where $[x]$ is largest possible integer not exceeding x. The square bracket is used in this sense throughout this paper.

The existence of the orthogonal array $(s^2, k, s, 2)$ is combinatorially equivalent to the existence of a set of $k - 2$ mutually orthogonal $s \times s$ Latin squares (such a set is usually said to have k constraints, represented by rows, columns and the $k - 2$ squares). The inequality (2.1) for the case $\lambda = 1$ states the well known fact that the maximum number of mutually orthogonal $s \times s$ Latin squares cannot exceed $s - 1$.

Again for the special case $s = 2$, (2.1) gives $k \leqq 4\lambda - 1$. Plackett and Burman have actually constructed orthogonal arrays $(4\lambda, 4\lambda - 1, 2, 2)$ for all values of $\lambda \leqq 25$, except for $\lambda = 23$. They also give a number of arrays of strength 2 for other values of s, and establish a connection between orthogonal arrays and affine resolvable balanced incomplete block designs [2], and between orthogonal arrays and partially balanced designs [3].

Rao [4] studies hypercubes of strength d, which are orthogonal arrays for which the index λ is a power of s. He has used them in connection with confounded factorial designs. The concept of orthogonal arrays in its most general form is also due to Rao [5]. He discusses the use of these arrays together with some methods of constructing them and gives the following generalization of the inequality of Plackett and Burman.

THEOREM. *For an orthogonal array $(\lambda s^t, k, s, t)$, $t \geqq 2$, the number of constraints k satisfies the inequality*

$$\lambda s^t - 1 \geqq C_1^k(s-1) + \cdots + C_u^k(s-1)^u \qquad \text{if } t = 2u, \tag{2.2}$$

$$\lambda s^t - 1 \geqq C_1^k(s-1) + \cdots + C_u^k(s-1)^u + C_u^{k-1}(s-1)^{u+1} \qquad \text{if } t = 2u+1. \tag{2.3}$$

When $t = 2$, this leads to Plackett and Burman's inequality (2.1). When $t = 3$, we get

COROLLARY. *For an orthogonal array $(\lambda s^3, k, s, 3)$ of strength 3, the number of constraints k satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{s - 1}\right] + 1. \tag{2.4}$$

Theorems 1A and 2A proved in Sections 3 and 4 give an alternative proof of the inequalities (2.1) and (2.4). Theorems 1B, 2B, 2C improve these inequalities except for certain special values of s.

Sections 5, 6 and 7 are devoted to the investigation of methods for constructing orthogonal arrays of strength 2. A difference theorem is proved, which when used in conjunction with Galois fields enables the construction of the arrays (18, 7, 3, 2) and (32, 9, 4, 2). The first of these has been constructed by Burman [1] by trial and error methods. It is shown that if p is prime and $s = p^v$, $\lambda = p^u$, $[u/v] = c$, then we can construct an orthogonal array $(\lambda s^2, k, s, 2)$, where $k = \{\lambda(s^{c+1} - 1)/(s^c - s^{c-1})\} + 1$.

Theorems 5A and 5B of Section 8 establish a connection between orthogonal arrays and the theory of confounding in symmetrical factorial designs (based on the use of finite projective geometries) first developed by Bose and Kishen [6] and later amplified by Bose [7]. It is shown that the problem of constructing the orthogonal array (s^r, k, s, t), $r \geq t$, $s = p^n$ and the problem of obtaining a symmetrical factorial design with s levels and k factors, in which the block size is s^r and in which all t-factor and lower order interactions are left unconfounded, both depend on finding a set of k points in $PG(r - 1, p^n)$ no t of which are conjoint. Such sets have been obtained by Bose in [7] and his results can be immediately translated into the language of orthogonal arrays. This has been done in Section 9. Theorems 5A and 5B were given by Bush in his unpublished thesis [8]. It has recently come to our notice that Rao [9] independently obtained a theorem equivalent to 5A, and derived from it the array $(2^r, 2^{r-1}, 2, 3)$ given by us in Section 9(a). The results in paragraphs (b) and (c) of Section 9 are new. An improvement of the inequalities (2.2) and (2.3) has been given by Bush [8, 10] for the special case $\lambda = 1$.

3. Upper bound for the number of constraints for orthogonal arrays of strength 2. Two columns of an orthogonal array are said to have i coincidences if there are exactly i rows in which the symbols appearing in the two columns have the same value (i.e., are the same elements of Σ). For example, the first column in the array (2.0) has 1 coincidence with each of the second and third columns, but has 3 coincidences with the fourth.

For any orthogonal array (N, k, s, t) of index λ let n_i denote the number of columns (other than the first) which have i coincidences with the first column. Since the total number of columns is $N = \lambda s^t$,

$$\sum_{i=0}^{k} n_i = \lambda s^t - 1. \tag{3.0a}$$

We shall show that

$$\sum_{i=0}^{k} i(i-1)\cdots(i-h+1)n_i = k(k-1)\cdots(k-h+1)(\lambda s^{t-h} - 1), \quad 1 \leq h \leq t. \tag{3.0b}$$

The formula (3.0a) can be regarded as a degenerate case of (3.0b) for $h = 0$.

Consider the subarray obtained by choosing any h rows of (N, k, s, t). The first column vector of this array appears in exactly $\lambda s^{t-h} - 1$ other columns of this subarray. Since it is possible to choose the subarray in C_h^k different ways the total number of $h \times 1$ vectors appearing in columns other than the first which are identical with the corresponding vector of the first column is $(\lambda s^{t-h} - 1)C_h^k$. But any column which has i-coincidences with the first contributes nothing or C_h^i to this number according as $i < h$ or $i \geq h$. Hence

$$\sum_{i=0}^{k} n_i C_h^i = C_h^k(\lambda s^{t-h} - 1), \tag{3.0c}$$

where C_h^i is to be interpreted as zero if $i < h$. This is equivalent to (3.0b).

Let us now confine our attention to orthogonal arrays of strength 2. Then (3.0a) and (3.0b) lead to

$$\sum_{i=0}^{k} n_i = \lambda s^2 - 1, \tag{3.1a}$$

$$\sum_{i=0}^{k} i n_i = k(\lambda s - 1), \tag{3.1b}$$

$$\sum_{i=0}^{k} i(i-1) n_i = k(k-1)(\lambda - 1). \tag{3.1c}$$

Consider the function

$$f(x) = \sum_{i=0}^{k} (i - x)(i - 1 - x) n_i,$$

defined for integral values of x. Then

$$f(x) \geqq 0 \tag{3.2}$$

since $n_i \geqq 0$, and the factors $(i - x)$ and $(i - 1 - x)$ are both negative if $i < x$, and both positive if $i > x + 1$. Also one factor is zero if $i = x$ or $x + 1$. Now

$$f(x) = \sum_{i=0}^{k} i(i-1)n_i - 2x \sum_{i=1}^{k} i n_i + x(x+1) \sum_{i=1}^{k} n_i;$$

whence from (3.1), we get

$$f(x) = \lambda\{k(k-1) - 2kxs + x(x+1)s^2\} - \{k(k-1) - 2kx + x(x+1)\}.$$

From (3.2)

$$\lambda \geqq \frac{k(k-1) - 2kx + x(x+1)}{k(k-1) - 2kxs + x(x+1)s^2}. \tag{3.3}$$

Setting

$$\alpha = k - 1 - xs,$$

we can after some reduction, write (3.3) in the form

$$\frac{\lambda s^2 - 1}{s - 1} \geqq k\left\{1 + \frac{\alpha(s - \alpha)}{D}\right\}, \tag{3.4}$$

where D can be expressed in two equivalent forms

$$D = (s-1)(k - \alpha - 1) + \alpha(\alpha + 1) \tag{3.5a}$$

$$= k(s-1) - (\alpha + 1)(s - 1 - \alpha). \tag{3.5b}$$

We shall now prove Plackett and Burman's inequality (2.1) for orthogonal arrays of strength 2, and then proceed to improve it if $\lambda - 1$ is not divisible by $s - 1$. Let

$$\lambda - 1 = a(s - 1) + b, \qquad 0 \leqq b < s - 1, \qquad a \geqq 0.$$

Therefore

$$\frac{\lambda s^2 - 1}{s - 1} = \lambda s + \lambda + a + \frac{b}{s - 1}. \tag{3.6}$$

Suppose there exists an array with $k = \lambda s + \lambda + a + 1$. Then

$$k - 1 = s(\lambda + a) + b + 1.$$

The integer x is at our disposal. Let us choose $x = \lambda + a$; then $\alpha = b + 1$, so that $0 < \alpha < s$.

From (3.5a) we have

$$D = s(s - 1)(\lambda + a) + \alpha(\alpha + 1) > 0;$$

so that

$$\frac{\alpha(s - \alpha)}{D} > 0.$$

Hence from (3.4) and (3.6)

$$\frac{b}{s - 1} > 1,$$

which is a contradiction. Hence, the value $k = \lambda s + \lambda + a + 1$ is inadmissible, and so are all higher values. This proves the inequality of Plackett and Burman.

THEOREM 1A. *For any orthogonal array* $(\lambda s^2, k, s, 2)$ *of strength* 2, *the number of constraints* k *satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{s - 1}\right].$$

Consider now the case when $\lambda - 1$ is not divisible by $s - 1$, so that $0 < b < s - 1$. Let

$$k = \lambda s + \lambda + a - n, \qquad b > n \geqq 0.$$

Therefore

$$k - 1 = s(\lambda + a) + b - n.$$

Choosing x as before, we now have

$$0 < \alpha = b - n < s - 1.$$

Therefore

$$(\alpha + 1)(s - 1 - \alpha) > 0.$$

Hence from (3.4) and (3.5b)

$$\frac{\lambda s^2 - 1}{s - 1} > k + \frac{\alpha(s - \alpha)}{(s - 1)},$$

or

$$\frac{b}{s-1} > -n + \frac{(b-n)(s-b+n)}{s-1}.$$

Therefore

$$(b-n)(b+1-n) - s(b-2n) > 0. \tag{3.7}$$

Hence if n is any integer $(b > n \geqq 0)$ for which the relation (3.7) is contradicted then the value $k = \lambda s + \lambda + a - n$ and all higher values are impossible. The first term in (3.7) is never negative, so that for $n > b/2$, this relation will never be contradicted. Hence we may drop the restriction $b > n$. The quadratic equation obtained by replacing the inequality by equality in (3.7) has one positive and one negative root, since the product of the roots is $-b(s-1-b)$ and $0 < b < s - 1$. The positive root may be written as

$$\theta = \frac{\sqrt{1 + 4s(s-1-b)} - (2s - 2b - 1)}{2}. \tag{3.8}$$

The largest value of n which contradicts (3.6) is $[\theta]$. Hence we may state the following theorem.

THEOREM 1B. *If* $\lambda - 1 = a(s-1) + b, 0 < b < s - 1$, *then for the orthogonal array* $(\lambda s^2, k, s, 2)$ *of strength* 2, *the number of constraints* k *satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{\lambda - 1}\right] - [\theta] - 1, \tag{3.9}$$

where θ *is the positive number given by* (3.8).

4. Upper bound for the number of constraints for orthogonal arrays of strength 3. Consider an array of strength 3, and let n_i denote the number of columns (other than the first) which have i coincidences with the first column. From (3.0a) and (3.0b)

$$\sum_{i=0}^{k} n_i = \lambda s^3 - 1, \tag{4.0a}$$

$$\sum_{i=0}^{k} i n_i = k(\lambda s^2 - 1), \tag{4.0b}$$

$$\sum_{i=0}^{k} i(i-1) n_i = k(k-1)(\lambda s - 1), \tag{4.0c}$$

$$\sum_{i=0}^{k} i(i-1)(i-2) n_i = k(k-1)(k-2)(\lambda - 1). \tag{4.0d}$$

If x is any positive integer then

$$f(x) = \sum_{i=0}^{k} i(i-1-x)(i-2-x) n_i \geqq 0; \tag{4.1}$$

whence from (4.0) we get

$$f(x) = k(k-1)(k-2)(\lambda - 1) - 2xk(k-1)(\lambda s - 1) + kx(x+1)(\lambda s^2 - 1) \geqq 0. \tag{4.2}$$

Since $k \geqq 1$, we have

$$\lambda \geqq \frac{(k-1)(k-2) - 2x(k-1) + x(x+1)}{(k-1)(k-2) - 2x(k-1)s + x(x+1)s^2}, \tag{4.3}$$

which is the same as (3.3) with $k - 1$ instead of k. Hence reasoning as before we can prove the following theorems.

THEOREM 2A. *For any orthogonal array* $(\lambda s^3, k, s, 3)$ *of strength* 3, *the number of constraints* k *satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{s - 1}\right] + 1. \tag{4.4}$$

THEOREM 2B. *If* $\lambda - 1 = a(s-1) + b, 0 < b < s - 1$, *then for the orthogonal array* $(\lambda s^3, k, s, 3)$ *of strength* 3, *the number of constraints* k *satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{\lambda - 1}\right] - [\theta], \tag{4.5}$$

where θ *is the positive number given by* (3.8).

Theorem 2A is the same as the Rao inequality (2.4) and Theorem 2B improves it for the case when $\lambda - 1$ is not divisible by $s - 1$.

We shall now show that when $\lambda - 1$ is divisible by $s - 1$, we can still improve the inequality of Theorem 2A, except in certain special cases. In fact we can state the following theorem.

THEOREM 2C. *For any orthogonal array* $(\lambda s^3, k, s, 3)$ *of strength* 3, *if* $\lambda - 1 = a(s-1)$, *and* $(s-1)^2(s-2)$ *is not divisible by* $as + 2$ *then the number of constraints* k *satisfies the inequality*

$$k \leqq \left[\frac{\lambda s^2 - 1}{s - 1}\right] - 1. \tag{4.6}$$

Now $[(\lambda s^2 - 1)/(s-1)] = as^2 + s + 1$. If possible let $k = as^2 + s + 1$. Choose $x = as$. Then it is easy to verify from (4.2) that $f(x) = 0$. Hence

$$\sum_{i=0}^{k} i(i - 1 - as)(i - 2 - as)n_i = 0.$$

Since $n_i \geqq 0$, it follows that n_i must vanish for all values of i except $i = 0, as + 1, as + 2$. From (4.0b) and (4.0c) we get

$$(as+1)n_{as+1} + (as+2)n_{as+2} = k(as^3 - as^2 + s^2 - 1),$$

$$asn_{as+1} + (as+2)n_{as+2} = ks(as^2 - as + s - 1).$$

Solving we get

$$n_{as+1} = k(s - 1),$$

$$n_{as+2} = \frac{ks(as^2 - 2as + a + s - 1)}{as + 2}$$

$$= k(s - 1)^2 - s(s - 1)(s - 2) + \frac{(s - 1)^2(s - 2)}{as + 2}.$$

Since n_{as+2} must be integral, we arrive at a contradiction if $(s - 1)^2(s - 2)$ is not divisible by $as + 2$. Hence in this case $k \leqq as^2 + s$.

Consider the special case $\lambda = s$. Then $a = 1$. If $(s - 1)^2(s - 2)/(s + 2)$ is integral, then 36 must be divisible by $s + 2$. We can therefore state the following corollary to Theorem 2C.

COROLLARY. *For the orthogonal array* $(s^4, k, s, 3)$ *if* 36 *is not divisible by* $s + 2$, *then the number of constraints* k *cannot exceed* $s^2 + s$.

5. The method of differences for constructing orthogonal arrays of strength 2. The method of differences has been elsewhere used [11] for constructing incomplete block designs. Here we shall use it to construct orthogonal arrays of strength 2.

Let $\lambda = \alpha\beta$. An orthogonal array $(\lambda s^2, k, s, 2)$ of strength 2 is said to be β-resolvable if it is the juxtaposition of $g = \alpha s$ different arrays $(\beta s, k, s, 1)$ of index β and strength 1. A 1-resolvable array is said to be completely resolvable. For example, the array (18, 6, 3, 2), obtained from (2.0) by deleting the last row is completely resolvable.

If $\lambda = \alpha\beta$ and the orthogonal array $(\lambda s^2, k, s, 2)$ is β-resolvable, then we can add at least one more row and get an orthogonal array of $k + 1$ constraints. In the new row we have to put the first element of Σ in the columns belonging to the first component array, the second element of Σ in the columns belonging to the next component and so on. As will be seen later under appropriate circumstances, it may be possible to add more than one row without destroying the orthogonality of the array.

THEOREM 3. *Let M be a module (additive group) consisting of s elements, e_0, $e_1, \cdots, e_{s-1}$. Suppose it is possible to find a scheme of r rows, with elements belonging to M*

$$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{matrix} \tag{5.0}$$

such that among the differences of the corresponding elements of any two rows, each element of M occurs exactly λ times ($n = \lambda s$); then the method of constructing a completely resolvable orthogonal array $(\lambda s^2, r, s, 2)$ of strength 2 is as follows: Write *down the addition table of M. Then replace each element in the scheme by the row of the addition table corresponding to the element (using only the suffixes if the set Σ is taken as $0, 1, \cdots, s-1$). This gives the completely resolvable array $(\lambda s^2, r, s, 2)$. A new row can be added to obtain an array $(\lambda s^2, r+1, s, 2)$ of $r+1$ constraints.*

Before proceeding to a formal proof we shall illustrate the use of the theorem, by constructing the orthogonal array (18, 7, 3, 2). For M we take the Galois field GF (3), whose elements are residue classes (mod 3). Let $e_0 = 0$, $e_1 = 1$, $e_2 = 2$. The addition table of M is

$$\begin{array}{c|ccc} & e_0 & e_1 & e_2 \\ \hline e_0 & e_0 & e_1 & e_2 \\ e_1 & e_1 & e_2 & e_0 \\ e_2 & e_2 & e_0 & e_1 \end{array} \tag{5.1}$$

It is not difficult to construct by trial a six rowed scheme

$$\begin{array}{cccccc} e_0 & e_0 & e_0 & e_0 & e_0 & e_0 \\ e_0 & e_0 & e_1 & e_2 & e_1 & e_2 \\ e_0 & e_1 & e_0 & e_2 & e_2 & e_1 \\ e_0 & e_2 & e_2 & e_0 & e_1 & e_1 \\ e_0 & e_1 & e_2 & e_1 & e_0 & e_2 \\ e_0 & e_2 & e_1 & e_1 & e_2 & e_0 \end{array} \tag{5.2}$$

where among the differences of the corresponding elements in any two rows each of the three elements e_0, e_1, e_2 occurs twice. In order to convert the scheme (5.2) into the completely resolvable orthogonal array (18, 6, 3, 2), we replace each element of M by the suffixes in the corresponding row of the addition table (5.1). Thus

$$e_0 \to 0,\ 1,\ 2,$$

$$e_1 \to 1,\ 2,\ 0,$$

$$e_2 \to 2,\ 0,\ 1.$$

We thus obtain the first six rows of the array (2.0) given in the Introduction.

Finally to obtain the array (18, 7, 3, 2) we add a new row consisting of six zeros (occupying the columns of the first two groups) followed by six ones, followed by six twos. It should be noted that from Theorem 1B, 7 is the maximum possible number of constraints for an array of size 18 and strength 2, with 3 levels.

We shall now proceed to a formal proof of Theorem 3. The s^2 2×1 vectors whose components are elements of M can be divided into s classes, each class corresponding to one element of M. If $e_i - e_j = e_k$ then $\begin{pmatrix} e_i \\ e_j \end{pmatrix}$ belongs to the class corresponding to e_k. Now in the addition table of M the difference of the corresponding elements of two different rows remains constant so that the vectors formed from the rows corresponding to e_i and e_j consist of all vectors of the class corresponding to e_k. Since in our scheme among the differences of corresponding elements of any two rows, each element of M occurs just λ times, when our scheme is expanded and each element replaced by the corresponding row of the addition table, every vector will occur λ times. (Replacing the elements by the corresponding suffixes will change the set Σ from M to the set $0, 1, 2, \cdots, s - 1$.)

6. Construction of a completely resolvable array $(\lambda s^2, \lambda s, s, 2)$ of strength 2 and λs constraints, when the index λ and the number of levels s are both powers of a prime p. Let $\lambda = p^u$, $s = p^v$. Consider the Galois field $GF(p^{u+v})$. The elements of the field can be expressed either as powers x^i of a primitive element $x(i = 0, 1, \cdots, p^{u+v} - 1)$ together with the element zero, or as polynomials of degree $u + v - 1$ with coefficient from $GF(p)$, the field of residue classes (mod p). (For a brief exposition of these properties of Galois fields see [11] and [12].) To add two elements we use the polynomial form adding the coefficients (mod p), and to multiply we use the power form remembering the relation

$$x^{p^{u+v}} = x. \tag{6.0}$$

For example if $p = 2$, $u = 1$, $v = 2$, we consider the Galois field $GF(2^3)$, whose elements may be exhibited (using the minimum function $x^3 + x^2 + 1$) as

$$\begin{aligned}
\alpha_0 &= 0 = 0\\
\alpha_1 &= 1 = x^0,\\
\alpha_2 &= x = x,\\
\alpha_3 &= x + 1 = x^5,\\
\alpha_4 &= x^2 = x^2,\\
\alpha_5 &= x^2 + 1 = x^3,\\
\alpha_6 &= x^2 + x = x^6,\\
\alpha_7 &= x^2 + x + 1 = x^4.
\end{aligned} \tag{6.1}$$

We have ordered the elements of the field in what may be called the lexicographic order, that is, if $\alpha_i = a_2x^2 + a_1x + a_0$ then the integer i is expressed as $a_2a_1a_0$ in the scale of numeration with radix 2. The same is done for the general case GF (p^{u+v}). If

$$\alpha_i = a_{n-1}x^{n-1} + \cdots + a_vx^v + a_{v-1}x^{v-1} + \cdots + a_1x + a_0, \tag{6.2}$$

then $i = a_{n-1} \cdots a_1a_0$ in the scale of numeration radix p, where $n = u + v$.

Consider the sub-class M of the elements of $GF(p^{u+v})$ for which the coefficients of x^v and higher powers of x are zero, when the element is expressed in the polynomial form. In our example the sub-class M consists of the elements α_0, α_1, α_2, α_3. In general M will consist of the first p^v elements of GF(p^{u+v}) when they are arranged in the lexicographic order. We now establish a correspondence between the elements of the field, and the elements of M in the following manner: The element α_i of GF(p^{u+v}) given by (6.2) corresponds to the element

$$\alpha_j = a_{v-1}x^{v-1} + \cdots + a_1x + a_0 \tag{6.3}$$

of M, the coefficients of x^{v-1} and lower powers of x for α_j being the same as the coefficients of the corresponding powers of x in α_i. It is clear that α_j is uniquely determined by α_i, and that

$$j = i \ (\text{mod } p^v), \qquad 0 \leq j < p^v.$$

Conversely to each α_j of M there correspond p^u elements of GF(p^{u+v}), since if α_j is given by (6.3) then for α_i the coefficients $a_{p-1}, \cdots, a_v$ are arbitrary each taking p possible values. It should be noticed that M is a direct factor module in GF(p^{u+v}) and that the correspondence used by us is a projection.

In the example under consideration the correspondence between the elements of GF(2^3) and M is given by

$$\begin{aligned} &\alpha_4, \alpha_0 \to \alpha_0, \\ &\alpha_5, \alpha_1 \to \alpha_1, \\ &\alpha_6, \alpha_2 \to \alpha_2, \\ &\alpha_7, \alpha_3 \to \alpha_3. \end{aligned} \tag{6.4}$$

If we write down the rows of the multiplication table of GF(p^{u+v}) and then replace each element by the corresponding element in M, we get a p^{u+v} rowed scheme which can be shown to satisfy the conditions of Theorem 3. If we take the difference of the corresponding elements in any two rows of the multiplication table, then every element of GF(p^{u+v}) occurs exactly once. Also if the elements α_i, $\alpha_{i'}$ of the field correspond to the elements of α_j, $\alpha_{j'}$ of M, then the element $\alpha_i - \alpha_{i'}$ of the field corresponds to the element $\alpha_j - \alpha_{j'}$ of M. This shows that in the scheme we have obtained each element of M occurs exactly $\lambda = p^u$ times, among the differences of the corresponding elements of any two rows. It follows from Theorem 3, that if each element of the scheme is now replaced by the corresponding row of the addition table of M (retaining only the suffixes) we get the completely resolvable array $(\lambda s^2, \lambda s, s, 2)$ where $\lambda = p^u$, $s = p^v$.

For example when $p = 2$, $u = 1$, $v = 2$, we have to write down the rows of the multiplication table of GF(2^3). This can be done by using the identifications given in (6.1), remembering that $x^8 = x$. We thus get

$$
\begin{array}{ll}
 & \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0 \\
 & \alpha_0\ \alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5\ \alpha_6\ \alpha_7 \\
 & \alpha_0\ \alpha_2\ \alpha_4\ \alpha_6\ \alpha_5\ \alpha_7\ \alpha_1\ \alpha_3 \\
(6.5) & \alpha_0\ \alpha_3\ \alpha_6\ \alpha_5\ \alpha_1\ \alpha_2\ \alpha_7\ \alpha_4 \\
 & \alpha_0\ \alpha_4\ \alpha_5\ \alpha_1\ \alpha_7\ \alpha_3\ \alpha_2\ \alpha_6 \\
 & \alpha_0\ \alpha_5\ \alpha_7\ \alpha_2\ \alpha_3\ \alpha_6\ \alpha_4\ \alpha_1 \\
 & \alpha_0\ \alpha_6\ \alpha_1\ \alpha_7\ \alpha_2\ \alpha_4\ \alpha_3\ \alpha_5 \\
 & \alpha_0\ \alpha_7\ \alpha_3\ \alpha_4\ \alpha_6\ \alpha_1\ \alpha_5\ \alpha_2 .
\end{array}
$$

Using the correspondence (6.4) the difference scheme is given by

$$
\begin{array}{ll}
 & \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0\ \alpha_0 \\
 & \alpha_0\ \alpha_1\ \alpha_2\ \alpha_3\ \alpha_0\ \alpha_1\ \alpha_2\ \alpha_3 \\
 & \alpha_0\ \alpha_2\ \alpha_0\ \alpha_2\ \alpha_1\ \alpha_3\ \alpha_1\ \alpha_3 \\
(6.6) & \alpha_0\ \alpha_3\ \alpha_2\ \alpha_1\ \alpha_1\ \alpha_2\ \alpha_3\ \alpha_0 \\
 & \alpha_0\ \alpha_0\ \alpha_1\ \alpha_1\ \alpha_3\ \alpha_3\ \alpha_2\ \alpha_2 \\
 & \alpha_0\ \alpha_1\ \alpha_3\ \alpha_2\ \alpha_3\ \alpha_2\ \alpha_0\ \alpha_1 \\
 & \alpha_0\ \alpha_2\ \alpha_1\ \alpha_3\ \alpha_2\ \alpha_0\ \alpha_3\ \alpha_1 \\
 & \alpha_0\ \alpha_3\ \alpha_3\ \alpha_0\ \alpha_2\ \alpha_1\ \alpha_1\ \alpha_2 .
\end{array}
$$

To obtain the completely resolvable array (32, 8, 4, 2) we replace the α's in (6.6) by the suffixes in the addition table of M. These replacements are

$$
\begin{array}{ll}
 & \alpha_0 \to 0,\ 1,\ 2,\ 3, \\
(6.7) & \alpha_1 \to 1,\ 0,\ 3,\ 2, \\
 & \alpha_2 \to 2,\ 3,\ 0,\ 1, \\
 & \alpha_3 \to 3,\ 2,\ 1,\ 0.
\end{array}
$$

Finally the orthogonal array (32, 9, 4, 2) can be obtained by adding a final row consisting successively of 8 zeros, 8 ones, 8 twos and 8 threes. The completed array is

$$
\begin{array}{ll}
 & 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3 \\
 & 0\ 1\ 2\ 3\ 1\ 0\ 3\ 2\ 2\ 3\ 0\ 1\ 3\ 2\ 1\ 0\ 0\ 1\ 2\ 3\ 1\ 0\ 3\ 2\ 2\ 3\ 0\ 1\ 3\ 2\ 1\ 0 \\
 & 0\ 1\ 2\ 3\ 2\ 3\ 0\ 1\ 0\ 1\ 2\ 3\ 2\ 3\ 0\ 1\ 1\ 0\ 3\ 2\ 3\ 2\ 1\ 0\ 1\ 0\ 3\ 2\ 3\ 2\ 1\ 0 \\
 & 0\ 1\ 2\ 3\ 3\ 2\ 1\ 0\ 2\ 3\ 0\ 1\ 1\ 0\ 3\ 2\ 1\ 0\ 3\ 2\ 2\ 3\ 0\ 1\ 3\ 2\ 1\ 0\ 0\ 1\ 2\ 3 \\
(6.9) & 0\ 1\ 2\ 3\ 0\ 1\ 2\ 3\ 1\ 0\ 3\ 2\ 1\ 0\ 3\ 2\ 3\ 2\ 1\ 0\ 3\ 2\ 1\ 0\ 2\ 3\ 0\ 1\ 2\ 3\ 0\ 1 \\
 & 0\ 1\ 2\ 3\ 1\ 0\ 3\ 2\ 3\ 2\ 1\ 0\ 2\ 3\ 0\ 1\ 3\ 2\ 1\ 0\ 2\ 3\ 0\ 1\ 0\ 1\ 2\ 3\ 1\ 0\ 3\ 2 \\
 & 0\ 1\ 2\ 3\ 2\ 3\ 0\ 1\ 1\ 0\ 3\ 2\ 3\ 2\ 1\ 0\ 2\ 3\ 0\ 1\ 0\ 1\ 2\ 3\ 3\ 2\ 1\ 0\ 1\ 0\ 3\ 2 \\
 & 0\ 1\ 2\ 3\ 3\ 2\ 1\ 0\ 3\ 2\ 1\ 0\ 0\ 1\ 2\ 3\ 2\ 3\ 0\ 1\ 1\ 0\ 3\ 2\ 1\ 0\ 3\ 2\ 2\ 3\ 0\ 1 \\
 & 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 3\ 3\ 3\ 3.
\end{array}
$$

It should be noted that 9 is the maximum possible number of constraints for an orthogonal array of size 32 and strength 2 with 4 levels (cf. Theorem 1B).

7. Adjunction of new rows to the completely resolvable array ($\lambda s^2, \lambda s, s, 2$), where $\lambda = p^u$, $s = p^v$ and p is a prime. As already explained we can add at least one more row to the array without destroying its orthogonality giving $\lambda s + 1$ constraints in all. Let

$$u = cv + d, \qquad c \geqq 0, \quad 0 \leqq d < v.$$

If $c = 0$, we stop after one row has been added. But if $c > 0$ we shall show that we can do better. Now $u \geqq v$. Let us denote by (A_0) the original completely resolvable array $(\lambda s^2, \lambda s, s, 2)$. Using the same construction as for (A_0), we can obtain another completely resolvable array $(\lambda_1 s^2, \lambda_1 s, s, 2)$ where $\lambda_1 = p^{u-v}$. Let us call this array (A_1). It should be noticed that the number of columns in (A_1) is equal to the number of arrays of strength unity composing (A_0) since $\lambda_1 s^2 = \lambda s = p^{u+v}$. We now inflate (A_1) by repeating each column s times, thus arriving at the array (A_1'), which has the same number of columns as (A_0). We now adjoin (A_1') to (A_0) placing the former just below (A_0). The result is that below any component of (A_0) we get the same column of (A_1) repeated s times. In view of the resolvability property of (A_0) it is clear that if we choose a particular row of (A_0) and a particular row of (A_1') then every ordered pair occurs λ times. Hence the whole array $\begin{pmatrix} A_0 \\ A_1' \end{pmatrix}$ is of strength 2 and has $\lambda s + \lambda_1 s$ constraints.

Since A_1 is completely resolvable, $\begin{pmatrix} A_0 \\ A_1' \end{pmatrix}$ is s-resolvable. If $c = 1$, then $\lambda_1 < s$, we stop after adjoining a final row to $\begin{pmatrix} A_0 \\ A_1' \end{pmatrix}$ consisting of λs zeros followed by λs ones and so on, getting $\lambda s + \lambda_1 s + 1$ constraints in all.

On the other hand if $c > 1$, we do not adjoin the final row as yet but construct a completely resolvable array $(\lambda_2 s^2, \lambda_2 s, s, 2)$ where $\lambda_2 = p^{u-2v}$. Denote this array by (A_2). We next inflate (A_2) to (A_2'') by repeating each column s^2 times and adjoin it to $\begin{pmatrix} A_0 \\ A_1' \end{pmatrix}$ arriving at the array $\begin{pmatrix} A_0 \\ A_1' \\ A_2'' \end{pmatrix}$ of strength 2 with $\lambda s + \lambda_1 s + \lambda_2 s$ constraints. If $c = 2$ we finish the process by adding the final row but if $c > 2$ we continue on as before.

The whole process therefore leads to an orthogonal array of strength 2 in which the number of constraints is given by

$$\lambda s + \lambda_1 s + \cdots + \lambda_c s + 1, \qquad \lambda_i = \lambda / s^i. \tag{7.0}$$

We can therefore state the following theorem.

THEOREM 4. *Given $s = p^v$, $\lambda = p^u$ (where p is a prime) then we can construct an orthogonal array $(\lambda s^2, k, s, 2)$ of strength 2, in which the number of constraints k is given by*

$$k = \frac{\lambda(s^{c+1} - 1)}{s^c - s^{c-1}} + 1, \tag{7.1}$$

where $c = [u/v]$.

8. The use of finite projective geometries in the construction of orthogonal arrays.

THEOREM 5A. *If we can find a matrix C of k rows and r columns*

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1r} \\ c_{21} & c_{22} & \cdots & c_{2r} \\ \vdots & \vdots & & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kr} \end{pmatrix} \tag{8.0}$$

whose elements c_{ij} belong to the Galois field $\mathrm{GF}(p^n)$, *and for which every partial matrix obtained by taking t rows is of rank t, then we can construct an orthogonal array (s^r, k, s, t), where $s = p^n$.*

PROOF. Consider $r \times 1$ column vectors ξ whose coordinates belong to $\mathrm{GF}(p^n)$. Then there are s^r different ξ. Form the matrix A whose s^r columns are the $k \times 1$ vectors $C\xi$. Then A is the required orthogonal array.

If A' is a $t \times s^r$ submatrix of A, and C' is the corresponding $t \times r$ submatrix of C, the columns α of A' are $C'\xi$, and since C' is of rank t, each α is obtained from s^{r-t} different ξ. Hence in A' each possible $t \times 1$ column vector occurs with the frequency $\lambda = s^{r-t}$, which shows that A is an orthogonal array of strength t and index λ.

The rows of the matrix C may be interpreted as the coordinates of a point in a finite projective space $\mathrm{PG}(r-1, p^n)$ such that no t of the points are conjoint. We thus get the following theorem:

THEOREM 5B. *If we can find k points in* $\mathrm{PG}(r-1, p^n)$ *so that no t are conjoint, then we can construct an orthogonal array (s^r, k, s, t) for which $\lambda = s^{r-t}$, $s = p^n$.*

It has been shown by Bose [7] that the maximum number of factors that it is possible to accommodate in a symmetrical factorial experiment in which each factor is at $s = p^n$ levels, and each block is of size s^r, without confounding any t-factor or lower order interaction, is given by the maximum number of points that it is possible to choose in the finite projective space $\mathrm{PG}(r-1, p^n)$ so that no t of the chosen points are conjoint (a set of t points are said to be conjoint if they lie on a flat space of dimensions not greater than $t-2$). This number is denoted by $m_t(r, s)$. It is clear from Theorem 5B that we can always construct an orthogonal array (s^r, k, s, t), for which the number of constraints $k \leqq m_t(r, s)$, if $s = p^n$ where p is a prime. The value of $m_t(r, s)$ has been determined by Bose in a number of important cases, and the corresponding set of points in which no t are conjoint has been obtained. These results are used in the next section to construct some orthogonal arrays of strength 3.

9. Construction of some orthogonal arrays of strength 3.

(a) Consider the special case $s = 2$. In $\mathrm{PG}(r-1, 2)$ consider the set of all points, which do not lie on the $(r-2)$-flat

$$x_1 + x_2 + \cdots + x_r = 0. \tag{9.0}$$

There are exactly 2^{r-1} such points, namely the points in whose coordinates there are an odd number of unities, and the rest zero. No three of these points are collinear since in PG$(r-1, 2)$ each line passes through exactly three points, and one of these lying in the plane (9.0) is excluded from our set. Taking the coordinates of these points for the rows of the matrix C of Theorem 5A, we can construct the orthogonal array $(2^r, 2^{r-1}, 2, 3)$ of strength 3 and 2^{r-1} constraints. Theorem 2A shows that this is the maximum possible number of constraints.

As an illustration consider the case $r = 3$. The four points of PG(2, 2) not lying on the line $x_1 + x_2 + x_3 = 0$ are (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1). Hence the corresponding matrix C is

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{9.1}$$

The eight possible column vectors ξ are

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1. \end{matrix}$$

The columns of the required array (8, 4, 2, 3) are obtained by forming $C\xi$ given below.

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1. \end{matrix} \tag{9.2}$$

Similarly the array (16, 8, 2, 3) is given by (9.3)

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1. \end{matrix} \tag{9.3}$$

(b) Let $s = 2^n$. In the finite projective plane PG$(2, 2^n)$ take the non-degenerate conic

$$ax_1^2 + bx_2^2 + cx_3^2 + fx_2x_3 + gx_3x_1 + hx_1x_2 = 0, \tag{9.4}$$

where

$$\Delta = af^2 + bg^2 + ch^2 + fgh \neq 0.$$

Of course no three of the points $P_1, P_2, \cdots, P_{s+1}$ on the conic (9.4) are collinear since no line can cut it in more than two points. Through any point (x_1', x_2', x_3') on the conic there pass $s + 1$ lines, of which s join it to the remaining points of the conic, while there is one line which does not meet the conic in any point other than (x_1', x_2', x_3'). This may be called the tangent at (x_1', x_2', x_3'). Its equation is

$$f(x_3'x_2 + x_2'x_3) + g(x_1'x_3 + x_3'x_1) + h(x_2'x_1 + x_1'x_2) = 0. \tag{9.5}$$

It is a peculiar feature of the finite projective geometry, based on a field of characteristic 2, that every tangent to a given non-degenerate conic passes through the same point. For example in the present case the arbitrary tangent (9.5), passes through the point P_0 with coordinates (f, g, h). The $s + 1$ tangents to (9.4) account for all the lines which pass through P_0. Hence no line through P_0 can meet the conic in more than one point. Thus $P_0, P_1, P_2, \cdots, P_{s+1}$ is a set of $s + 2$ points, such that no three are collinear. Hence from Theorem 5B we can use the coordinates of these points to construct an orthogonal array $(s^3, s + 2, s, 3)$ where $s = 2^n$.

Similarly when $s = p^n$ where p is an odd prime we could construct the array $(s^3, s + 1, s, 3)$ by using the coordinates of the $s + 1$ points on a non-degenerate conic of $PG(2, p^n)$.

One of the authors, Bush [10], has shown that for an orthogonal array (s^t, k, s, t) of index unity and strength t, the number of constraints k satisfies the inequality

$$k \leqq s + t - 1 \qquad \text{when } s \text{ is even,} \tag{9.6a}$$

$$k \leqq s + t - 2 \qquad \text{when } s \text{ is odd.} \tag{9.6b}$$

Using this result for $t = 3$, we find that the number of constraints obtained by us for arrays of size s^3, s levels and strength 3, cannot be improved.

(c) Let $\phi(x, y) = ax_1^2 + 2hx_1x_2 + bx_2^2$ be a homogeneous quadratic with coefficients belonging to $GF(p^n)$ and irreducible in it. If $s = p^n$, it can be shown [7] that the quadratic surface

$$ax_1^2 + 2hx_1x_2 + bx_2^2 = x_3x_4$$

contains exactly $s^2 + 1$ points no three of which are collinear. We therefore get a method of constructing an orthogonal array $(s^4, k, s, 3)$ with $k = s^2 + 1$ constraints, when s is a prime or a prime power. On the other hand, Theorem 2C gives an upper bound $s^2 + s$ for k when $s \neq 2, 4, 7$ or 16 and an upper bound for k for these exceptional values of s is given as $s^2 + s + 2$ by Theorem 2A. Thus there remains a gap between the number of constraints which might be attainable, and the number of constraints actually attained except for the case $s = 2$, for which we have already obtained an array $(s^4, 8, 2, 3)$ by the method (a). It is not known whether this gap can be bridged. It has been shown [7] that when p is odd we cannot get more than $s^2 + 1$ points in $PG(3, p^n)$ no three of which are collinear. The same has been proved by Seiden [13] for the case $s = 2^2$.

Hence for these cases the geometrical method cannot lead to more than $s^2 + 1$ constraints, but there remains the possibility that some other combinatorial procedure may lead to a larger number of constraints.

As example consider the case $s = 3$. The coordinates of the 10 points lying on the quadric $x_1^2 + x_2^2 = x_3x_4$ of PG(3, 3) are (0, 0, 1, 0), (0, 0, 0, 1), (0, 1, 1, 1), (0, 1, 2, 2), (1, 0, 1, 1), (1, 0, 2, 2), (1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 2), (1, 2, 2, 1). Using these as the rows of the matrix C, we get the orthogonal array (81, 10, 3, 3), and 10 is the maximum number of constraints obtainable by the geometrical methods. Theorem 2C gives $k \leqq 12$. We do not know whether we can get 11 or 12 constraints in any other way.

REFERENCES

[1] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiments," *Biometrika*, Vol. 33 (1943–1946), pp. 305–325.
[2] R. C. Bose, "A note on the resolvability of balanced incomplete block designs," *Sankhyā*, Vol. 6 (1942), pp. 105–110.
[3] R. C. Bose and K. R. Nair, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337–373.
[4] C. R. Rao, "Hypercubes of strength 'd' leading to confounded designs in factorial experiments," *Bull. Calcutta Math. Soc.*, Vol. 38 (1946), pp. 67–78.
[5] C. R. Rao, "Factorial experiments derivable from combinatorial arrangements of arrays," *Jour. Royal Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 128–139.
[6] R. C. Bose and K. Kishen, "On the problem of confounding in the general symmetrical factorial design," *Sankhyā*, Vol. 5 (1940), pp. 21–36.
[7] R. C. Bose, "Mathematical theory of symmetrical factorial designs," *Sankhyā*, Vol. 8 (1947), pp. 107–166.
[8] K. A. Bush, "Orthogonal Arrays," unpublished thesis, University of North Carolina (1950).
[9] C. R. Rao, "On a class of arrangements," *Edinburgh Math Soc.*, Vol. 8 (1949), pp. 119–125.
[10] K. A. Bush, "Orthogonal arrays of index unity," *Annals of Math Stat.*, Vol. 23 (1952), pp. 426–434.
[11] R. C. Bose, "On the construction of balanced incomplete block designs," *Annals of Eugenics*, Vol. 9 (1939), pp. 353–399.
[12] R. C. Bose, "On the application of the properties of Galois fields to the construction of hyper Graeco-Latin squares," *Sankhyā*, Vol. 3 (1938), pp. 323–338.
[13] E. Seiden, "A theorem in finite projective geometry and an application to statistics," *Proc. Amer. Math Soc.*, Vol. 1 (1950), pp. 282–286.

A NONPARAMETRIC TEST FOR THE SEVERAL SAMPLE PROBLEM[1]

BY WILLIAM H. KRUSKAL

University of Chicago

1. Summary. Suppose that C independent random samples of sizes $n_1, \cdots, n_C$ are to be drawn from C univariate populations with unknown cumulative distribution functions $F_1, \cdots, F_C$. This paper discusses a test of the null hypothesis $F_1 = F_2 = \cdots = F_C$ against alternatives of the form

$$F_i(x) = F(x-\theta_i) \qquad (\text{all } x, i = 1, \cdots, C)$$

with the θ_i's not all equal, or against alternatives of a much more general sort to be specified in Section 5. The test to be discussed has as its critical region large values of the ordinary F-ratio for one-way analysis of variance, computed after the observations have been replaced by their ranks in the $\sum n_i$-fold over-all sample. This use of ranks simplifies the distribution theory, and permits application of the test to cases where the ranks are available but the numerical values of the observations are difficult to obtain. Briefly, then, we shall consider a nonparametric analogue, based on ranks, of one-way analysis of variance.

It is shown in Section 4 that, under quite general conditions, the proposed test statistic, H, is asymptotically chi-square with $C - 1$ degrees of freedom when the null hypothesis holds. Section 5 derives a necessary and sufficient condition that the natural family of sequences of tests based on large values of H all be consistent against a given alternative. Section 6 derives the variance of H under the null hypothesis, Section 7 derives the maximum value of H, and Section 8 gives a difference equation which may be used to obtain exact small-sample distributions under the null hypothesis. These derivations are made on the assumption of continuity for the cumulative distribution functions; Section 9 considers extensions to the possibly discontinuous case.

2. Introduction. Until Section 9 all cumulative distribution functions will be supposed continuous. The over-all sample consists of the $\sum n_i = N$ (say) independent random variables $\xi_j^{(i)}$ (i = 1, $\cdots$, C; $j = 1, \cdots, n_i$), where the superscript refers to the (sub)sample and the subscript indexes observations within a (sub)sample. Under the null hypothesis all the ξ's have the same continuous but unknown cdf (cumulative distribution function): $F(x)$. Each $\xi_j^{(i)}$ is immediately replaced by $X_j^{(i)}$, its rank in the over-all sample. Then, under the null hypothesis, the N-tuple $(X_1^{(1)}, \cdots, X_{n_1}^{(1)}, X_1^{(2)}, \cdots, X_{n_2}^{(2)}, \cdots\cdots, X_1^{(C)}, \cdots, X_{n_C}^{(C)})$ takes as values with equal probability the $N!$ permutations of $(1, 2, \cdots, N)$.

Next let $R_i = \sum_{j=1}^{n_i} X_j^{(i)}$ be the sum of ranks of sample from the ith population and let $\bar{R}_i = R_i/n_i$. Of course $\sum R_i = \frac{1}{2}N(N+1)$. The standard one-way

[1] Work done under the sponsorship of the Office of Naval Research.

analysis of variance test based on the X's has for its critical region large values of

$$\sum_{i=1}^{C} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2 \Big/ \left[\sum_{i=1}^{C}\sum_{j=1}^{n_i}(X_j^{(i)} - \bar{R}_i)^2\right].$$

But this is a monotone increasing function of

$$\sum_{i=1}^{C} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2 \Big/ \left[\sum_{j=1}^{C}\sum_{j=1}^{n_i}\left(X_j^{(i)} - \frac{N+1}{2}\right)^2\right] \tag{2.1}$$

and because of the use of ranks the denominator of the above expression is a constant. Hence a critical region consisting of large values of the numerator of (2.1) is suggested. The corresponding test is the one to be discussed in this paper. Actually this test will be discussed in terms of the random variable

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{C} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2 = \frac{12}{N(N+1)}\sum_{i=1}^{C}\frac{R_i^2}{n_i} - 3(N+1).$$

Since the variance of the uniform distribution on the integers $1, 2, \cdots, N$ is $(N^2-1)/12$, it is natural to expect that the numerator of the F-ratio in terms of ranks divided by this variance is asymptotically chi-square with $C-1$ degrees of freedom. But this normalized numerator is just $H\ N/(N-1)$. The minor advantage of H over this and other asymptotically equivalent random variables upon which the test might be based is that under the null hypothesis $EH = C - 1 = E$(chi-square with $C - 1$ degrees of freedom).

3. Relationship to other tests. When $C = 2$ the test discussed in this paper is the same as the symmetrical two-tail version of a test considered by Wilcoxon [11] and by Mann and Whitney [2]; for when $C = 2$

$$H = \frac{12}{(N+1)n_1(N-n_1)}\left(R_1 - n_1\frac{N+1}{2}\right)^2.$$

For $C = 2$, a test against *any* alternative (subject to existence of and weak conditions on a density function) is provided by the work of Wald and Wolfowitz [9] who propose and discuss a test based on runs in the over-all sample. A generalization of the Wald-Wolfowitz test for any C is available (e.g., Wallis [12]). For any C a test based on the median of the over-all sample and reducing to a conventional chi-square test has been proposed by Brown and Mood in Chapter 16 of [4]. A generalization of this test using several previously determined order statistics of the over-all sample is described by Massey [3]. Other tests are discussed in [1] and [13]. For $C = 2$ a recent addition to the list of tests has been made by Marshall [14].

Whitney [10] in the case of $C = 3$ considers two tests designed to have particular power against the following two types of alternatives, respectively:

(1) $F_1(x) > F_2(x)$ and $F_1(x) > F_3(x)$ (all x),

(2) $F_1(x) > F_2(x) > F_3(x)$ (all x).

His tests appear to be generalizable for any C.

The test discussed in this paper, that is, that based on large values of H, is closely akin to tests considered by Pitman, Friedman and others for two-way analysis of variance problems (see the expository paper by Scheffé [7] for a general discussion and references). However, this particular application of the randomization method has not to my knowledge been discussed in the statistical literature. Further discussion of related tests will be found in [15].

4. Asymptotic distributions. The term "asymptotic distribution" will be taken in the sense of convergence in distribution as $N \to \infty$. We shall assume at present that, for all i, $\lim_{N\to\infty} n_i/N = \nu_i$ exists and is positive. We proceed to show that under the null hypothesis the R_i's, properly normed, have asymptotically a singular multivariate normal distribution, and that from this the asymptotic chi-square distribution of H readily follows. The proof will be a direct application of a powerful general theorem of Wald and Wolfowitz [8], and I shall suppose that the reader is familiar with this reference. A consequence of the Wald-Wolfowitz theorem, in a form appropriate for our purpose, may be stated as follows:

THEOREM 4W (*Wald-Wolfowitz*). *Let* $\{A_N\}$ *be a sequence of ordered N-tuples,* $A_N = (a_{N1}, a_{N2}, \cdots, a_{NN})$, $(N = 1, 2, 3, \cdots)$ *satisfying condition W of* [8]. *Let* $(Z_{N1}, \cdots, Z_{NN})$ *for each N be a random ordered N-tuple taking as values with equal probability the permutations of* A_N. *Let* $\{n_i^{(N)}\}$ $(i = 1, \cdots, C; N = 1, 2, 3, \cdots)$ *be C sequences of non-negative integers such that*

$$\sum_{i=1}^{C} n_i^{(N)} = N, \qquad \lim_{N\to\infty} n_i^{(N)}/N = \nu_i$$

exists. Let $L_N^{(i)} = \sum Z_{N\alpha}$ *for* $i = 1, \cdots, C$ *where the summation is from* $\alpha = \sum_{j=1}^{i-1} n_j^{(N)} + 1$ *to* $\alpha = \sum_{j=1}^{i} n_j^{(N)}$. *Let* v_N *be the common variance of any* $Z_{N\alpha}$. *Then, asymptotically, the random variables* $[L_N^{(i)} - EL_N^{(i)}]/\sqrt{Nv_N}$ *have the singular C-variate normal distribution with mean zero and covariance matrix whose i, i′ term is*

$$\delta_{ii'}\,\nu_i - \nu_i\,\nu_{i'}. \tag{4.1}$$

The proof of Theorem 4W follows from Corollary 1 of [8] via the technique used in Section 7 of [8], that is, via the consideration of arbitrary linear combinations of the random variables $[L_N^{(i)} - EL_N^{(i)}]/\sqrt{Nv_N}$. In order to save space the details are omitted. Note that for Theorem 4W itself it is not necessary to assume that the ν_i's are positive.

To apply Theorem 4W to our case, set $a_{N\alpha} = \alpha$ and observe that the resulting $\{A_N\}$ satisfies condition W of [8] (see, e.g., Section 3 of [8]). $L_N^{(i)}$ is called R_i, and it may be readily computed that $ER_i = \frac{1}{2} n_i^{(N)} (N + 1)$ and $v_N = (N^2 - 1)/12$. Hence the variables

$$\sqrt{12}\,\frac{R_i - n_i\dfrac{N+1}{2}}{N^{3/2}}$$

(dropping the superscript "N" for convenience) have asymptotically the singular multivariate normal distribution with zero mean and covariance matrix given by (4.1). We next use the assumption that all the ν_i's are positive, from which it follows immediately that the variables

$$T_i = \sqrt{12}\,\frac{R_i - ER_i}{N^{3/2}\sqrt{n_i/N}}$$

have a joint asymptotic normal distribution with zero mean vector and with the covariance matrix whose i, i' term is $\delta_{ii'} - \sqrt{\nu_i \nu_{i'}}$. We now make the standard analysis of variance transformation

$$S_0 = \sum_{i'=1}^{C} \sqrt{\nu_{i'}}\, T_{i'}, \qquad S_i = \sum_{i'=1}^{C} e_{ii'} T_{i'} \qquad (i = 1, 2, \cdots, C-1)$$

with the e's chosen to make the transformation orthogonal. It follows that $\sum_{i=1}^{C} T_i^2$ is asymptotically chi-square with $C-1$ degrees of freedom. But

$$\sum_{i=1}^{C} T_i^2 = \frac{12}{N^2}\sum_{i=1}^{C}\frac{1}{n_i}\left(R_i - n_i\frac{N+1}{2}\right)^2 = \frac{N+1}{N}H.$$

Hence H is asymptotically chi-square with $C-1$ degrees of freedom.

It seems desirable to make a few comments regarding possible weakening of the conditions for an asymptotic chi-square distribution. In the first place, no great difficulty arises if some ν_i's should be zero—for example, suppose that $\nu_1 = 0$ and the other ν_i's are positive. Then $(R_1 - ER_1)/N^{3/2}$ approaches zero stochastically and $[N/(N+1)] \sum_{i=2}^{C} T_i^2$, that is, H computed from the sample *without* including R_1, is asymptotically chi-square with $C-2$ degrees of freedom. It is not however true in general that $\sum_{i=1}^{C} T_i^2$ is asymptotically chi-square; for example, consider the case of $n_1 = 1$ for all N. Analogous remarks apply if more than one $\nu_i = 0$.

If we use chi-square with $C-1$ degrees of freedom to approximate the critical region, then it may be wise to drop from the total sample any (sub)samples with n_i's very small. We would do this in order to obtain a better approximation to the critical region, at the expense of losing power against certain alternatives involving the populations from which the omitted observations arose. On the other hand, because of the smallness of the n_i's in question, even the exact critical region would probably have had little power against these alternatives.

Whitney in [10] uses a kind of limit requirement which might be thought applicable here and weaker than the existence of the ν_i's; that is to suppose the existence, for $i \neq j$, of

$$\tau_{ij} = \lim_{N\to\infty} \frac{n_i n_j}{(N - n_i)(N - n_j)}.$$

(Assume all $n_i < N$, so that the τ's are defined.) That this requirement is little weaker than ours except effectively for the case $C = 2$ is shown by the following lemma.

LEMMA 4.1. *If the ν_i's exist and no $\nu_i = 1$, then the τ_{ij}'s exist; and if $\nu_i = 0$, then every $\tau_{ij} = 0$ ($j = 1, 2, \cdots, C, j \neq i$). When $C \geq 3$ if the τ_{ij}'s exist and at* *least two different τ_{ij}'s are $\neq 0$, then the ν_i's exist; and if $\tau_{ij} = 0$, then either ν_i or $\nu_j = 0$.*

PROOF. The first part is obvious. To prove the second choose an i, say $i = 1$ for convenience. Suppose that we can then find a j and k ($j \neq k; j, k \neq 1$) such that $\tau_{jk} \neq 0$. Then

$$\tau_{1j}\,\tau_{1k}/\tau_{jk} = \lim_{N\to\infty}\left(\frac{n_1}{N - n_1}\right)^2.$$

Hence ν_1 exists, and if $\tau_{1j} = 0$ then $\nu_1 = 0$. Next suppose that no such τ_{jk} exists, that is, that the only possible nonzero τ's are $\tau_{12}, \tau_{13}, \cdots, \tau_{1C}$. By hypothesis at least two of these must be nonzero, say τ_{12} and τ_{13}. Since $\tau_{23} = 0$, it follows as above that ν_2 and ν_3 exist and are zero. The same comment holds for any other ν_j for which $\tau_{1j} \neq 0$. Finally suppose a $\tau_{1j} = 0$ ($j \neq 2, 3$). Since $\tau_{12} \neq 0$ we have

$$\lim_{N\to\infty}\frac{n_j/(N - n_j)}{n_2/(N - n_2)} = 0.$$

But the denominator here approaches zero itself. Hence the limit of $n_j/(N-n_j)$ is zero, ν_j exists, and it is zero. Of course $\nu_1 = 1$.

If only one τ, say τ_{12}, is $\neq 0$, then $\nu_3, \nu_4, \cdots, \nu_C$ must all be 0. It can be shown that $\tau_{12} = 1$, as well, so that we are effectively in the $C = 2$ case. It is impossible for all the τ's to be zero.

The material of this section is summarized in the following theorem:

THEOREM 4. *If for all i, lim $n_i/N = \nu_i$ exists and is positive, then under the null hypothesis H is asymptotically distributed as chi-square with $C - 1$ degrees of freedom. If p ν_i's are zero ($p = 1, 2, \cdots, C - 1$), then H computed with only the R_i's corresponding to nonzero ν_i's is asymptotically distributed under the null hypothesis as chi-square with $C - p - 1$ degrees of freedom.*

5. Consistency of the test based on large values of H. Suppose that the n_i's are functions of N, and consider the family of sequences of critical regions $H \geqq t_\alpha(N)$, where the level of significance $\alpha \,\varepsilon\, (0, 1)$ indexes the sequences of the family, N indexes the members of a sequence, and $t_\alpha(N)$ is the least number with the property $Pr\{H \geqq t_\alpha(N)\} \leqq \alpha$ under the null hypothesis. Let us say that this family of sequences is consistent against a given alternative if every member sequence is consistent in the usual sense against the given alternative, that is, if for all $\alpha \,\varepsilon\, (0, 1)$

$$\lim_{N\to\infty} \Pr\{H \geqq t_\alpha(N)\} = 1,$$

where the probabilities are taken under the alternative. For brevity we may simply say that the test based on large values of H is consistent. (Note that failure of consistency for the family of sequences against an alternative implies only that there is some α_0 such that for all $\alpha \leqq \alpha_0$ the sequence of tests $H \geqq t_\alpha(N)$ fails to be consistent in the usual sense.) This use of the word "consistent" will permit more compact statements and will not, I think, cause any confusion.

Under what circumstances is the test based on large values of H consistent? We consider alternatives of the following form: all the ξ's are independent, and $\xi_j^{(i)}$ has a continuous cdf F_i. Assume that as $N \rightarrow \infty$, $n_i/N = \nu_i + o(N^{-\frac{1}{2}})$ with $\nu_i > 0$; and note that this assumption subsumes the most natural case: $n_i = [\nu_i N]$ or $[\nu_i N] + 1$. Since $t_\alpha(N)$ for given α has as its limit the upper 100α-percent point of the chi-square distribution with C-1 degrees of freedom, it is equivalent to ask under what circumstances $\lim_{N\rightarrow\infty} \Pr\{H \geqq t\} = 1$ for all positive t. We may also replace H by $\sum T_i^2$ since the two differ by a factor of $N/(N+1)$.

First, we ask under what circumstances, for all positive t, $\lim_{N\rightarrow\infty} \Pr\{| T_1 | \geqq t\} = 1$. Following the useful procedure of Mann and Whitney, set V = the number of couples $(X_j^{(1)}, X_{j'}^{(i)})$ where $i = 2$ or 3 or $\cdots C$ and for which $X_j^{(i)} > X_{j'}^{(i)}$. Then

$$R_1 = \tfrac{1}{2} n_1(n_1 + 1) + V.$$

This relationship holds for the special case $X_j^{(1)} \leqq n_1, j = 1, 2, \cdots, n_1$; for then $V = 0$. It holds in general, since an interchange of the superscripts of two adjacent X's, one from sample 1 and the other from sample $i \neq 1$, increases or decreases R_1 and V together by unity. Then

$$\begin{aligned} T_1 &= \sqrt{\frac{12}{n_1 N^2}} \left\{ \frac{n_1(n_1+1)}{2} + V - n_1 \frac{N+1}{2} \right\} \\ &= \sqrt{\frac{12}{n_1 N^2}} \{V - \tfrac{1}{2} n_1 (N - n_1)\}. \end{aligned}$$

Next define the following $n_1(N-n_1)$ counter-variables

$$Y_{jj'}^{(i)} = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} \text{ when } X_j^{(1)} \begin{Bmatrix} \leqq \\ > \end{Bmatrix} X_{j'}^{(i)}$$

so that

$$V = \sum_{i=2}^{C} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_i} Y_{jj'}^{(i)}.$$

From now on we deal with a specific alternative F_i which will be described in slightly different terms further on.

LEMMA 5.1. *The set of values of Var T_1, as N runs through the positive integers, is bounded.*

PROOF. Set Var $Y_{j_1 j_2}^{(i)} = v_i$ and

$$\text{Cov}\,(Y_{j_1 j_2}^{(i)}, Y_{j_3 j_4}^{(i')}) = \begin{cases} c_{ii'} \text{ for } j_1 = j_3, \text{ and either } i \neq i' \text{ or } j_2 \neq j_4, \\ d_i \text{ for } i = i', j_2 = j_4, \text{ and } j_1 \neq j_3, \\ 0 \text{ otherwise (since } \xi\text{'s independent).} \end{cases}$$

Clearly the v's, c's, and d's are all less than 1 absolutely. So

$$\operatorname{Var} T_1 = \frac{12}{N^2 n_1} \operatorname{Var} V$$

$$\leqq \{\text{number of } v_i \text{ terms} + \text{number of } c_{ii'} \text{ terms} + \text{number of } d_i \text{ terms}\}$$

$$\leqq \frac{12}{N^2 n_1} \{n_1(N - n_1) + n_1(N - n_1)^2 + n_1^2(N - n_1)\} = \frac{12(N - n_1)(N + 1)}{N^2}$$

which is finite and has the limit $12(1 - \nu_1) < \infty$. We next introduce the numbers

$$g_{i,i'} = \Pr\{X_j^{(i)} > X_{j'}^{(i')}\}$$

(under the alternative) for $i \neq i'$, and $g_{i,i'} = \frac{1}{2}$. Hence for $i > 1$ $g_{1,i} = EY_{jj'}^{(i)}$ and

$$ET_1 = \sqrt{\frac{12}{n_1 N^2}} \left[n_1 \sum_{i=2}^{c} n_i g_{1,i} - \tfrac{1}{2} n_1(N - n_1) \right].$$

Hence the limit of $ET_1/\sqrt{n_1}$ is

$$\sqrt{12} \left[\sum_{i=2}^{c} \nu_i g_{1,i} - \tfrac{1}{2}(1 - \nu_1) \right] = \sqrt{12} \left[\sum_{i=1}^{c} \nu_i g_{1,i} - \tfrac{1}{2} \right],$$

whence the limit of ET_1 is

$$\left.\begin{matrix} \infty \\ 0 \\ -\infty \end{matrix}\right\} \text{as } \sum_{i=1}^{c} \nu_i g_{1,i} \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} \tfrac{1}{2},$$

and we have

LEMMA 5.2. *If* $\sum_{i=1}^{c} \nu_i g_{1,i} \neq \frac{1}{2}$, *then* $\lim_{N\to\infty} \Pr\{| T_1 | \geqq t\} = 1$. This follows immediately from Tchebycheff's inequality. Consequently we may state

LEMMA 5.3. *If for some* i, $\sum_{i'=1}^{c} \nu_{i'} g_{i,i'} \neq \frac{1}{2}$, *then the test based on large values of* H *is consistent.* We now turn to implications in the other direction.

LEMMA 5.4. *If* $\sum_{i=1}^{c} \nu_i g_{1,i} = \frac{1}{2}$, *then there exists a* t_0, *a function* $N_0(t)$, *and a decreasing function* $G(t)$ *such that*

(1) $\lim_{t\to\infty} G(t) = 0$,

(2) *For* $t > t_0$ *and* $N > N_0(t)$, $\Pr\{| T_1 | \geqq t\} \leqq G(t) < 1$.

PROOF. Let K be an upper bound for $\operatorname{Var} T_1$. Then by Tchebycheff's inequality, for $t > 0$

$$\Pr\{\, | T_1 | \geqq t\} = \Pr\{T_1 \geqq t\} + \Pr\{T_1 \leqq -t\} \leqq \frac{K}{[t - ET_1]^2} + \frac{K}{[t + ET_1]^2}$$

which has the limit $2K/t^2$, since $ET_1 \to 0$. Putting $\epsilon(t) = [\max(1, t)]^{-1}/4$ and $t_0 = 2\sqrt{K}$, it follows that for any $t > t_0$, there is an $N_0(t)$ such that for $N > N_0(t)$

$$\Pr\{\, | T_1 | \geqq t\} \leqq \frac{2K}{t^2} + \epsilon(t) < \tfrac{1}{2} + \tfrac{1}{4} < 1.$$

$G(t)$ is the function in the middle of the above double inequality. Next this is generalized as follows:

LEMMA 5.5. *If for all* i, $\sum_{i'=1}^{C} \nu_{i'} g_{i,i'} = \frac{1}{2}$, *then the test based on large values of* H *is not consistent.*

PROOF. By the previous lemma there are (for each i) numbers, $t_0^{(i)}$, and functions $N_0^{(i)}(t)$ and $G^{(i)}(t)$ such that $G^{(i)}(t) \to 0$ from above monotonically as $t \to \infty$, and such that for $t > t_0^{(i)}$ and $N > N_0^{(i)}(t)$, $Pr\{|T_i| \geqq t\} \leqq G^{(i)}(t) < 1$. Let $t_0^* = \max_i t_0^{(i)}$, $N_0^*(t) = \max_i N_0^{(i)}(t)$, and $G^*(t) = \max_i G^{(i)}(t)$. Then $G^*(t)$ is a monotone decreasing function with limit 0, and for all i, $t > t_0^*$, and $N > N_0^*(t)$, $\Pr\{| T_i | \geqq t\} \leqq G^*(t) < 1$. For $t > t_0^*$ and $N > N_0^*(t)$, $\Pr\{\text{some } | T_i | \geqq t\} \leqq C \cdot G^*(t)$. But $\sum T_i^2 \geqq s > 0$ implies that some $| T_i | \geqq \sqrt{s/C}$. Hence for $s > C \cdot t_0^{*2}$, and $N > N_0^* (\sqrt{s/C})$ we have

$$\Pr\{\sum T_i^2 \geqq s\} \leqq C \cdot G^* (\sqrt{s/C})$$

To complete the proof take s large enough so that $C \cdot G^* (\sqrt{s/C}) < 1$.

It is natural to ask for a simple probabilistic interpretation of the necessary and sufficient condition for consistency which has been proven. This may be done as follows. Recall that we are still discussing a fixed alternative $\{F_i\}$ and that all probabilities are taken with respect to this alternative. Now let $\eta^{(1)}, \cdots, \eta^{(C)}$ be C independent random variables independent of all the ξ's and with cdf's F_i. Then

$$g_{i,i'} = \Pr\{\eta^{(i)} > \xi_j^{(i')}\}.$$

Next choose a $\xi_J^{(I)}$ at random from among the N possibilities (i.e., take an observation in the space of N ordered couples (I, J) where each has the same probability $1/N$.) Then

$$\Pr \{\eta^{(i)} > \xi_J^{(I)}\} = \frac{1}{N} \sum_{i'=1}^{C} \sum_{j=1}^{n_{i'}} g_{i,i'} = \sum_{i'=1}^{C} \frac{n_{i'}}{N} g_{i,i'}$$

so that the test based on large values of H is inconsistent if and only if for all i, $\lim_{N\to\infty} \Pr\{\eta^{(i)} > \xi_J^{(I)}\} = \frac{1}{2}$. Roughly speaking this means that the test is consistent if and only if the variables from at least one population tend in the limit to be either larger or smaller than the other variables.

In particular we have consistency under the following circumstances which generalize to the C-population case the sufficient conditions for $C = 2$ given in [2] by Mann and Whitney[2]

$$F_1(x) < F_2(x), F_1(x) \leqq F_i(x) \qquad (i = 3, 4, \cdots, C)$$

for all x. (Of course the choice of subscripts 1 and 2 here is just for convenience.) To show that the consistency condition is satisfied, note that for $i = 3, 4, \cdots, C$

[2] The unnecessary specialization of the Mann and Whitney consistency condition when $C = 2$ was noted (separately) by Lehmann and van Dantzig; see p. 166 of [1] and [16]. In the latter both sufficiency and necessity are considered by a method similar to that of this paper, and further results are obtained. In 1948 E. J. G. Pitman gave the same necessary and sufficient condition for $C = 2$ during lectures at Columbia University.

$$g_{1,i} = \int_{-\infty}^{\infty} F_i(x)\, dF_1(x) \geqq \int_{-\infty}^{\infty} F_1(x)\, dF_1(x) = \tfrac{1}{2}$$

because of symmetry. However $g_{1,2} > \frac{1}{2}$; for let m run through the positive integers and set B_m equal to the set of all x satisfying $F_1(x) + 1/m \geqq F_2(x) > F_1(x) + 1/(m+1)$. Then

$$\begin{aligned} g_{1,2} &= \int_{-\infty}^{\infty} F_2(x)\, dF_1(x) = \sum_{m=1}^{\infty} \int_{B_m} F_2(x)\, dF_1(x) \\ &\geqq \sum_{m=1}^{\infty} \left\{ \int_{B_m} F_1(x)\, dF_1(x) + \frac{1}{m+1} \int_{B_m} dF_1(x) \right\} \\ &= \frac{1}{2} + \sum_{m=1}^{\infty} \frac{1}{m+1} \int_{B_m} dF_1(x), \end{aligned}$$

and clearly at least one B_m must have positive measure with respect to F_1. Hence $\sum_{i=1}^{C} \nu_i g_{1,i} > \frac{1}{2}$, and we have consistency. The circumstances just discussed include the translational sort of alternative described in the introductory paragraph of this paper.

A simple class of cases for which consistency fails and yet the null hypothesis need not hold is given by the following characteristic: that all the C distributions be symmetrical about the same point f in the following sense:

$$F_i(f-x) = 1 - F_i(f+x)$$

for all i and x. For, setting $f = 0$ without loss of generality, this means that the distribution of every ξ is the same as that of its negative. Hence for all i, i', $g_{i,i'} = g_{i',i} = \frac{1}{2}$ and consistency fails.

The material of this section may be summarized as follows.

THEOREM 5. *Suppose that the n_i's are functions of N and that for all i, $n_i/N = \nu_i + o(N^{-\frac{1}{2}})$ and $\nu_i > 0$. For each level of significance $\alpha (0 < \alpha < 1)$ consider the sequence of tests: reject the null hypothesis if $H \geqq t_\alpha(N)$ where $t_\alpha(N)$ is the least number giving rise to level of significance α at the Nth step. Then these sequences of tests are all consistent against a given continuous alternative $\{F_i\}$ if and only if for some i, with probabilities taken under the alternative*

$$\sum_{i'=1}^{C} \nu_{i'} [\Pr\{\eta^{(i)} > \eta^{(i')}\} + \tfrac{1}{2} \Pr\{\eta^{(i)} = \eta^{(i')}\}] \neq \tfrac{1}{2},$$

where the $\eta^{(i)}$'s are C independent random variables having respectively the cdf's F_i. The sufficiency of the above condition holds regardless of the order of $(n_i/N) - \nu_i$. When $C = 2$ the denial of the above condition implies $g_{12} = g_{21} = \frac{1}{2}$.

6. The variance of H under the null hypothesis. As an aid in approximating the distribution of H when the null hypothesis is true, we seek the variance of H under the null hypothesis. This seems to be a tedious computation by any method; we shall outline a direct method, omitting most of the routine algebra. Directly from the definition of H we have

$$\text{(6.1)} \quad \operatorname{Var} H = \frac{144}{N^2(N+1)^2} \left\{ \sum_i \frac{1}{n_i^2} ER_i^4 + \sum_{i \neq j} \frac{1}{n_i n_j} E(R_i^2 R_j^2) - \left[\sum_i \frac{1}{n_i} ER_i^2 \right]^2 \right\}.$$

$E\,R_i^4$ is readily found from formula (8) of [2], which when translated into our notation says

$$E\left(R_i - n_i\,\frac{N+1}{2}\right)^4 = \frac{n_i(N-n_i)(N+1)}{240}$$
$$[5Nn_i(N-n_i) - 2n_i^2 - 2(N-n_i)^2 + 3n_i(N-n_i) - 2N].$$

From this

$$\frac{1}{n_i^2}\,ER_i^4 = \frac{N+1}{240}\left\{n_i^2[15N^3 + 15N^2 - 10N - 8] + n_i\,[30N^3 + 50N^2 + 16N] + [5N^3 + 9N^2 + 2N] - \frac{1}{n_i}\,[2N^3 + 2N^2]\right\},$$

and, summing over i

$$\sum_i \frac{1}{n_i^2}\,ER_i^4 = \frac{N+1}{240}\,[15N^3 + 15N^2 - 10N - 8]\sum_i n_i^2$$
$$+\frac{N^2(N+1)}{240}\,[30N^2 + 50N + 16] + \frac{C(N+1)N}{240}\,[5N^2 + 9N + 2] \tag{6.2}$$
$$-\frac{N+1}{240}\,[2N^3 + 2N^2]\sum_i \frac{1}{n_i}.$$

Next we find $E(R_1^2R_2^2)$ as follows:

$$E(R_1^2R_2^2) = \sum_i\sum_{i'}\sum_j\sum_{j'} E[X_i^{(1)}\,X_{i'}^{(1)}\,X_j^{(2)}\,X_{j'}^{(2)}],$$

where $i, i' = 1, 2, \cdots, n_1$ and $j, j' = 1, 2, \cdots, n_2$. This quantity is

$$n_1n_2(n_1-1)(n_2-1)E[X_1^{(1)}\,X_2^{(1)}\,X_1^{(2)}\,X_2^{(2)}] + n_1n_2(n_1-1)E[X_1^{(1)}\,X_2^{(1)}\,X_1^{(2)2}]$$
$$+ n_1n_2(n_2-1)\,E\,[X_1^{(1)2}X_1^{(2)}\,X_2^{(2)}] + n_1n_2E\,[X_1^{(1)2}\,X_1^{(2)2}]$$
$$= \frac{n_1n_2(n_1-1)\,(n_2-1)}{N(N-1)(N-2(N-3)}\sum pp'\,qq' + \frac{n_1n_2(n_1+n_2-2)}{N(N-1)(N-2)}\sum p^2qq'$$
$$+\frac{n_1n_2}{N(N-1)}\sum p^2q^2,$$

where the p's and q's run from 1 to N and within any term of a summation no two are equal. This simplifies after some algebraic labor; we divide by n_1n_2 and make the obvious generalization from (1, 2) to (i, j) to find

$$\frac{1}{n_in_j}\,E(R_i^2R_j^2) = (n_i-1)(n_j-1)\,\frac{N+1}{240}\,[15N^3 + 15N^2 - 10N - 8]$$
$$+ (n_i + n_j - 2)\,\frac{N+1}{360}\,[30N^3 + 35N^2 - 11N - 12]$$
$$+\frac{N+1}{180}\,[20N^3 + 24N^2 - 5N - 6].$$

Sum this over $i \neq j$ to obtain

$$\begin{aligned}&\frac{N+1}{240}[(N-C)^2+2N-C-\sum_i n_i^2][15N^3+15N^2-10N-8]\\ (6.3)\qquad &+\frac{N+1}{360}[2(C-1)(N-C)][30N^3+35N^2-11N-12]\\ &+\frac{N+1}{180}C(C-1)[20N^3+24N^2-5N-6].\end{aligned}$$

Next, we note that

$$\frac{1}{n_i}ER_i^2=\frac{N+1}{12}(N-n_i)+n_i\frac{(N+1)^2}{4}=\frac{N+1}{12}[n_i(3N+2)+N]$$

which, summed over i and squared, gives

$$(6.4)\qquad \frac{N^2(N+1)^2}{144}[9N^2+6N(C+2)+(C+2)^2].$$

Finally, substituting (6.2), (6.3), and (6.4) in (6.1), and simplifying, we obtain

$$(6.5)\qquad \begin{aligned}&\text{Var } H\\ &=2(C-1)-\frac{2}{5N(N+1)}[3C^2-6C+N(2C^2-6C+1)]-\frac{6}{5}\sum_i\frac{1}{n_i}.\end{aligned}$$

Note that as all the n_i's $\to \infty$, Var $H \to 2(C-1) =$ Var (chi-square with $C-1$ degrees of freedom).

7. The maximum value of H. It is an aid in approximating the distribution of H to know its maximum value. This may be obtained from the well-known analysis of variance algebraic identity

$$(7.1)\qquad \begin{aligned}\frac{N(N+1)}{12}H+\sum_{i=1}^{C}\sum_{j=1}^{n_i}(X_j^{(i)}-\bar{R}_i)^2&=\sum_{i=1}^{C}\sum_{j=1}^{n_i}\left(X_j^{(i)}-\frac{N+1}{2}\right)^2\\ &=\frac{1}{12}N(N^2-1).\end{aligned}$$

A sample point maximizing H is a sample point minimizing the second term on the left side of the above identity, that is the within sum of squares. Clearly this sum of squares is minimized when the ranks within each (sub) sample form consecutive integers, that is $X_j^{(i)}-\bar{R}_i=j-\frac{1}{2}(n_i-1)$, so that

$$\begin{aligned}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(X_j^{(i)}-\bar{R}_i)^2&=\frac{1}{12}\sum_{i=1}^{C}n_i(n_i^2-1)\\ &=\frac{1}{12}(\sum n_i^3-N).\end{aligned}$$

Substituting back in (7.1) it follows that the maximum value of H is

$$(7.2)\qquad (N-1)+\frac{N-\sum n_i^3}{N(N+1)}=\frac{N^3-\sum n_i^3}{N(N+1)}.$$

8. On the distribution of the R_i's and H. If we set

$$\Gamma(r_1, r_2, \cdots, r_C; n_1, n_2, \cdots, n_C)$$

$$= \frac{\left(\sum_{j=1}^{C} n_j\right)!}{n_1! \cdots n_C!} Pr\{R_i = r_i \quad (i = 1, 2, \cdots, C) \text{ with stated } n_i\text{'s}\},$$

then, under the null hypothesis, the following difference equation holds for Γ.

$$\Gamma(r_1, \cdots, r_C; n_1, \cdots, n_C) \tag{8.1}$$

$$= \sum_{i=1}^{C} \Gamma\left(r_1, \cdots, r_{i-1}, r_i - \sum_{j=1}^{C} n_j, r_{i+1}, \cdots, r_C; n_1, \cdots, n_{i-1}, n_i - 1, n_{i+1}, \cdots, n_C\right)$$

with the following boundary conditions:

1. If any argument fails to be a non-negative integer, $\Gamma = 0$.
2. If $r_i n_i = 0$, but $r_i + n_i \neq 0$, then $\Gamma = 0$.
3. When $n_1 = n_2 = \cdots = n_C = 0$, $\Gamma = 1$ when all r_i's are 0, and otherwise $\Gamma = 0$.

The above equation and conditions follow readily from the partition of the chance event $\{R_1 = r_1, \cdots, R_C = r_C\}$ into the C chance events $\{R_1 = r_1, \cdots, R_C = r_C$, and $\max_{i',j} X_j^{(i')}$ is an $X_j^{(i)}\}$ for $i = 1, 2, \cdots, C$.

I have been unable to find a closed solution for this equation. For $C = 2$ and small n_i's, values of $[n_1!n_2!/(n_1+n_2)!]\ \Gamma(r_1, r_2; n_1, n_2)$ are given in [2]. (Comment on notation: m, n, and U of [2] correspond to our n_2, n_1, and $R_1 - \frac{1}{2}\ n_1(n_1+1)$ respectively. The tables of [2] actually give the cumulative probabilities.) For $C = 3$ and for small values of the n_i's, recursive computations based on (8.1) are being carried out in order to obtain exact distributions of H. It is hoped that from these exact distributions some idea of the accuracy of various approximations may be obtained.

In Section 6 formula (6.5) for the variance of H under the null hypothesis was obtained as a function of N, C, and $\sum 1/n_i$. It seems reasonable to attempt to better the chi-square approximation by fitting a Type III distribution with density function

$$\frac{a^\nu}{\Gamma(\nu)} t^{\nu-1} e^{-at}$$

for $t \geqq 0$, and 0 for $t < 0$. Equating first and second moments

$$a = (C-1)/\text{Var}\, H, \qquad \nu = (C-1)^2/\text{Var}\, H.$$

Equivalently, one may approximate $Pr\{H \leqq x\}$ under the null hypothesis by the use of K. Pearson's incomplete Γ function tables [5], setting u and p of those tables equal respectively to

$$x/\sqrt{\text{Var}\, H}, \qquad (C-1)^2/\text{Var}\, H - 1.$$

Again equivalently, one can make the same approximation by interpolation in a chi-square table using $2(C-1)^2/\text{Var } H$ degrees of freedom and argument $2x(C-1)/\text{Var } H$.

In Section 7 formula (7.2) for the maximum value of H, say H_M, was obtained in terms of N and $\sum n_i^3$. It seems reasonable to attempt to better the chi-square and Type III approximations by fitting an incomplete B distribution and using [6] to obtain approximate probabilities. Thus, equating moments again, we may approximate $\Pr\{H \leqq x\}$ under the null hypothesis by $I_{x/H_M}(p, q)$, in the notation of [6], where

$$p = \frac{C-1}{H_M} \cdot \frac{(C-1)(H_M - C + 1) - \text{Var } H}{\text{Var } H}, \qquad q = \frac{H_M - C + 1}{C-1}\, p.$$

The above formulas are given for convenient reference. The relative merits of these approximations will be discussed in [15].

9. The possibly discontinuous case. Much of the preceding material carries almost directly over to the general case in which the cdf's need not be continuous, providing that the following randomization convention is followed: when two or more ξ's are equal, define the corresponding X's at random. More precisely, suppose that $\xi_{j_1}^{(i_1)} = \xi_{j_2}^{(i_2)} = \cdots = \xi_{j_\omega}^{(i_\omega)}$ for a given sample point, and that all other ξ's are unequal to the common value of the above ω ξ's with (say) λ ξ's less than the common value. Then assign ranks $\lambda+1, \lambda+2, \cdots, \lambda+\omega$ to the tied ξ's by performing a random experiment in which each of the $\omega!$ possible assignments is an equally likely outcome. With this convention the joint distribution of the X's under the null hypothesis is the same as that stated in Section 2, so that the asymptotic chi-square distribution (Section 4) holds.

The following minor changes would be made in Section 5:

(1) In the discussion of the intuitive interpretation for the consistency condition, replace the given expression for $g_{i,i'}$ by $g_{i,i'} = \Pr\{\eta^{(i)} > \xi_j^{(i')}\} + \frac{1}{2}\Pr\{\eta^{(i)} = \xi_j^{(i')}\}$, and replace the necessary and sufficient condition for inconsistency by

$$\lim_{N\to\infty} \left[\Pr\{\eta^{(i)} > \xi_j^{(I)}\} + \tfrac{1}{2}\Pr\{\eta^{(i)} = \xi_j^{(I)}\}\right] = \tfrac{1}{2}, \text{ for all } i.$$

(2) In the discussion of consistency when $F_1 < F_2$, and $F_1 \leqq F_i$ $(i = 3, 4, \cdots, C)$, insert the remark that the result continues to hold if we consider the cdf's not in one of the usual senses (i.e., continuous to the left or to the right), but rather in the sense of Lévy: $\frac{1}{2}F(x^-) + \frac{1}{2}F(x^+)$. The same interpretation of the cdf notation should be made in the discussion of a class of cases for which consistency fails.

(3) Delete the word "continuous" in the statement of Theorem 5.

Another way to treat ties, much discussed in connection with the rank correlation coefficient, is to give tied ξ's equal fractional ranks so as to keep the sum of ranks at its usual value; i.e., in the notation of the first paragraph of this section, assign the fractional rank $\lambda + \frac{1}{2}(\omega + 1)$ to all the ω tied ξ's. We proceed to show that if we do this, and also change the norming constants appropriately, the altered H still is asymptotically chi-square with $C - 1$ degrees of freedom.

Suppose that there are K groups of ties with, respectively, $\omega_1, \omega_2, \cdots, \omega_K$ members. We agree to use mean ranks in the tied groups and to work in the conditional distribution wherein just K tied groups exist of sizes $\omega_1, \cdots, \omega_K$ and covering fixed rank intervals, but permitting the numbers of observations from the C subsamples falling in any tied group to vary. In other words, instead of the finite population $(1, 2, \cdots, N)$, we deal with

$$
\begin{array}{lll}
 & 1, 2, \cdots, \lambda_1, & \\
 & \lambda_1 + \frac{1}{2}(\omega_1 + 1), \cdots, \lambda_1 + \frac{1}{2}(\omega_1 + 1), & \lambda_1 + \omega_1 + 1, \cdots, \lambda_2, \\
(9.1) & \lambda_2 + \frac{1}{2}(\omega_2 + 1), \cdots, \lambda_2 + \frac{1}{2}(\omega_2 + 1), & \lambda_2 + \omega_2 + 1, \cdots, \lambda_3, \\
 & \vdots & \\
 & \lambda_K + \frac{1}{2}(\omega_K + 1), \cdots, \lambda_K + \frac{1}{2}(\omega_K + 1), & \lambda_K + \omega_K + 1, \cdots, N,
\end{array}
$$

where $\lambda_k + \frac{1}{2}(\omega_k + 1)$ occurs ω_k times. Under the null hypothesis the ordered N-tuple of $X_j^{(i)}$'s takes as its values the permutations of the above finite population, all with equal probability. We compute that $ER_i = \frac{1}{2}n_i(N + 1)$, as before, and that

$$
\operatorname{Var} R_i - \frac{n_i(N - n_i)(N + 1)}{12} = -\frac{n_i(N - n_i)}{N(N - 1)} \sum_{k=1}^{K} \frac{1}{12}\, \omega_k(\omega_k - 1)(\omega_k + 1),
$$

$$
\operatorname{Cov}(R_i, R_{i'}) + \frac{n_i n_{i'}(N + 1)}{12} = \frac{n_i n_{i'}}{N(N - 1)} \sum_{k=1}^{K} \frac{1}{12}\, \omega_k(\omega_k - 1)(\omega_k + 1),
$$

so that, setting

$$
\gamma = \sum_{k=1}^{K} \omega_k(\omega_k - 1)(\omega_k + 1),
$$

we have

$$
E[(R_i - ER_i)(R_{i'} - ER_{i'})] = \frac{1}{12}[n_i N \delta_{ii'} - n_i n_{i'}] \frac{N^3 - N - \gamma}{N(N - 1)}
$$

or the corresponding second moment in the untied case times

$$
[N^3 - N - \gamma]/[N^3 - N].
$$

Now let the λ_k's, the ω_k's and the n_i's all be functions of N, and assume that $\lim_{N\to\infty} n_i/N = \nu_i > 0$ exists, $\lim_{N\to\infty} \gamma/N^3 = \gamma^*$ exists and $\gamma^* \neq 1$. To say that $\gamma^* \neq 1$ is to say that $\operatorname{Max}_k \omega_k/N$ does not approach 1, and one can readily show then that the sequence of finite populations (9.1) satisfies condition W of Theorem 4W. It follows from Theorem 4W that the variables

$$
\sqrt{12}\, \frac{R_i - n_i \dfrac{N + 1}{2}}{\sqrt{N^3 - N - \gamma}}
$$

are asymptotically multivariate normal with zero mean and covariance matrix given by (4.1). Hence, just as in Section 4

$$\frac{12}{N(N+1)\left[1-\frac{\gamma}{N(N^2-1)}\right]}\sum_{i=1}^{c}\frac{1}{n_i}\left(R_i-n_i\frac{N+1}{2}\right)^2=H^* \tag{9.2}$$

(say) is asymptotically chi-square with C − 1 degrees of freedom and has expected value C − 1. Note that no limit condition on the λ_k's is needed.

10. Further work. It would be interesting to investigate further the power function of the test described in this paper, perhaps along the lines of [1], or by considering its asymptotic relative efficiency to ordinary one-way analysis of variance in the normal case. Again in the spirit of [1], it would seem desirable to propose and investigate related tests specifically designed to be powerful against more restricted classes of alternatives, e.g., $F_1 \geqq F_2 \geqq \cdots \geqq F_C$, with at least one inequality strong.[3] Another extension is to consider the use of H-like tests in two-way analyses of variance or more general linear hypothesis situations, in a manner analogous to that of [4].

11. Acknowledgments. The test discussed here was suggested to me in a slightly variant form by W. A. Wallis. I wish to acknowledge with gratitude Professor Wallis' encouragement and helpful suggestions in carrying through the work reported here. In particular Professor Wallis suggested the applicability of the mean rank approach in the case of ties and obtained the proper norming constant. I wish also to thank D. A. S. Fraser for his derivation, replacing a longer version, of the maximum value of H.

REFERENCES

[1] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests" *Annals of Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[2] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[3] F. J. Massey, Jr., "A note on a two sample test," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 304–306.

[4] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill Book Co., New York, 1950.

[5] K. Pearson, *Tables of the Incomplete Γ-Function*, His Majesty's Stationery Office, London, 1922.

[6] K. Pearson, *Tables of the Incomplete* B*-function*, Cambridge University Press, 1934.

[7] H. Scheffé, "Statistical inference in the non-parametric case," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 305–332.

[8] A. Wald and J. Wolfowitz, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 358–372.

[9] A. Wald and J. Wolfowitz, "On a test whether two samples are from the same population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 147–162.

[10] D. R. Whitney, "A bivariate extension of the U statistic," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 274–282.

---

[3] A recent investigation in this direction is that of T. J. Terpstra [17].

[11] F. Wilcoxon, "Individual comparison by ranking methods," *Biometrics Bull.*, Vol. 1 (1945), p. 80–83.

[12] W. A. Wallis, "Rough and ready statistical tests," *Industrial Quality Control*, Vol. 8 (1952), p. 35–40.

[13] J. Wolfowitz, "Non-parametric statistical inference," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949.

[14] A. W. Marshall, "A large sample test of the hypothesis that one of two random variables is stochastically larger than the other," *Jour. Am. Stat. Assn.*, Vol. 46 (1951), pp. 366–374.

[15] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion analysis of variance," *Jour. Am. Stat. Assn.*, Vol. 47 (1952), pp. 583–621.

[16] D. Van Dantzig, "On the consistency and power of Wilcoxon's two sample test," *Indagationes Mathematicae*, Vol. 13 (1951), pp. 1–8.

[17] T. J. Terpstra, "The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking," *Indagationes Mathematicae*, Vol. 14 (1952), pp. 327–333.

TESTING MULTIPARAMETER HYPOTHESES[1]

By E. L. Lehmann

Stanford University and University of California, Berkeley

1. Summary. Let the distribution of some random variables depend on real parameters $\theta_1, \cdots, \theta_s$ and consider the hypothesis $H: \theta_i \leqq \theta_i^*, i = 1, \cdots, s$. It is shown under certain regularity assumptions that unbiased tests of H do not exist. Tests of minimum bias and other types of minimax tests are derived under suitable monotonicity conditions. Certain related multidecision problems are discussed and two-sided hypotheses are considered very briefly.

2. Introduction. The extensive literature on optimum tests has been concerned mainly with hypotheses specifying a set of values for a single real valued parameter. Important exceptions are some cases that can be reduced to the one-parameter situation by the principle of invariance, such as the linear (univariate) hypothesis and Hotelling's T^2-problem. These have been used to illustrate a number of different principles, the successful application of which however seems to rest on the symmetry whose full exploitation makes the problems uniparametric. Another exception is the theory of tests with local optimum properties, initiated by Neyman and Pearson [1] and recently developed further by Isaacson [2].

We shall here concern ourselves mainly with hypotheses which, rather than specifying the values of the parameter in question, state that these parameters do not exceed certain bounds. The following examples illustrate the way in which such problems arise.

Example 2.1. Let p and p' denote the number of major and minor defects in a lot. Then the lot will be considered acceptable provided $p \leqq p_0$ and $p' \leqq p_0'$, where $p_0 < p_0'$.

Example 2.2. It may be desired to compare some new treatments with a standard one. Here the hypothesis would specify that none of the new treatments is better by more than a given amount than the standard.

Example 2.3. Let $x_1, \cdots, x_n$ be a sample from a normal distribution with mean ξ and variance σ^2. The population in question may be considered adequate if $\xi \leqq \xi_0$ and $\sigma \leqq \sigma_0$.

In some of the above examples we are dealing with *bona fide* testing problems while in others we are faced with a choice among more than two decisions. Which of these is the case cannot always be seen from the mathematical formulation alone. Thus in Example 2.1 it clearly depends on the disposition that is made of a rejected lot. If there is complete screening, the reason for rejection is immaterial. If on the other hand a lot rejected for major defects is treated differently from one rejected only for minor defects the decision problem becomes more complicated.

[1] Work done under the sponsorship of the Office of Naval Research.

We shall in the following concern ourselves mainly with the one-sided case of the straightforward testing problem. The two-sided situation and the multi-decision problem will be discussed only rather briefly. For simplicity we shall take the number of parameters to be two. The extension to the higher dimensional cases is immediate.

3. Unbiasedness and the minimax principle. The success of the concept of unbiasedness in the one-parameter case suggests the use of this approach also for the present problems. Unfortunately it turns out that in general unbiased tests of the hypotheses in question do not exist. Let us consider the case of two parameters θ_1, θ_2 and the hypothesis $H:\theta_1 \leqq \theta_1^*, \theta_2 \leqq \theta_2^*$. We shall assume that the power function $\beta(\theta_1, \theta_2)$ of any test is analytic in θ_1 and θ_2 in the sense that it can be expanded in an absolutely convergent double power series. Then we shall show that for any unbiased test we have $\beta(\theta_1, \theta_2) \equiv \alpha$, so that any unbiased test is equivalent to the trivial one that rejects with probability α regardless of the observations. This incidentally proves this trivial and most unsatisfactory test to be admissible for the problem under consideration.

Without loss of generality assume that $\theta_1^* = \theta_2^* = 0$. Then unbiasedness states that $\beta(\theta_1, \theta_2) \leqq \alpha$ in the third quadrant of the θ_1, θ_2-plane, and $\geqq \alpha$ in the other three quadrants. By continuity we have $\beta(\theta_1, 0) = \alpha$ for $\theta_1 \leqq 0$ and hence by analyticity $\beta(\theta_1, 0) = \alpha$ for all θ_1. Analogously $\beta(0, \theta_2) \equiv \alpha$. Consider now $\beta(\theta_1, \theta_2)$ for any fixed $\theta_2 > 0$ as a function of θ_1. It has a minimum at $\theta_1 = 0$ so that $\partial\beta(\theta_1, \theta_2)/\partial\theta_1 \mid_{\theta_1=0} = 0$ for all $\theta_2 \geqq 0$. Since $\partial\beta(\theta_1, \theta_2)/\partial\theta_1 \mid_{\theta_1=0}$ is again analytic in θ_2, it follows that $(\partial\beta(\theta_1, \theta_2/\partial\theta_1) \mid_{\theta_1=0}$ is identically zero. Consider now $\beta(\theta_1, \theta_2)$ for some fixed value $\theta_2 \leqq 0$. Since $\beta(\theta_1, \theta_2) \gtrless \alpha$ as $\theta_1 \gtrless 0$ and since at $\theta_1 = 0$ the derivative is zero, $\beta(\theta_1, \theta_2)$ must have a point of inflection at 0 and consequently the second derivative $\partial^2\beta(\theta_1, \theta_2)/\partial\theta_1^2 \mid_{\theta_1=0} = 0$ for all $\theta_2 < 0$ and hence for all θ_2. Since the, order of the first non-vanishing derivative $\partial^k\beta(\theta_1, \theta_2)/\partial\theta_1^k \mid_{\theta_1=0}$ must be even for $\theta_2 > 0$ and odd for $\theta_2 < 0$ we see in this manner that for any fixed θ_2 $\partial\, \beta(\theta_1, \theta_2)^k/\partial\theta_1^k \mid_{\theta_1=0} = 0$ for all $k = 1, 2, \cdots$. By analyticity it follows for each fixed θ_2 that $\beta(\theta_1, \theta_2)$ must be a constant, that is, be independent of θ_1. By symmetry it now follows that $\beta(\theta_1, \theta_2)$ must be identically constant, as was to be proved.

We digress for a moment from our search for a reasonable test of the hypothesis $\theta_1, \theta_2 \leqq 0$ to point out that there do exist non-trivial tests of H satisfying the condition of similarity

$$\beta(\theta_1, 0) = \beta(0, \theta_2) = \alpha$$

for all θ_1, θ_2. Suppose for example that X and Y are independently distributed with joint density $f_{\theta_1}(x)f_{\theta_2}(y)$ and that $\alpha = 1/m$ where m is an integer. Then we can obtain a particularly simple class of similar regions as follows. Let $S_1, \cdots, S_m$ be mutually exclusive and exhaustive sets on the real line such that

$$\int_{S_i} f_0(x)\, dx = \alpha, \qquad i = 1, \cdots, m.$$

In the x, y-plane define the set w_i to be the Cartesian product of S_i with itself, and let $w = w_1 + \cdots + w_m$. Now when $\theta_2 = 0$, X is a sufficient statistic for θ_1 and for every x we have

$$P_{\theta_1,0}((X, Y) \varepsilon W \mid x) = \alpha,$$

and hence $P_{\theta_1,0}((X, Y) \varepsilon W) \equiv \alpha$.

We now return to the original problem, and investigate the test that maximizes the minimum power over a certain class ω' of alternatives. For ω' we take the set of points (θ_1, θ_2) for which either $\theta_1 \geqq \theta_1^{**}$ or $\theta_2 \geqq \theta_2^{**}$. Let us consider first as an example the case that X and Y are independently, normally distributed, with known variances and with means θ_1 and θ_2, respectively. Then it would seem as if any reasonable test should satisfy the following two conditions:

(i) $\beta(\theta_1, \theta_2) \leqq \beta(\theta_1', \theta_2')$ whenever $\theta_1 \leqq \theta_1'$, $\theta_2 \leqq \theta_2'$, if ϕ denotes the critical function,

(ii) $\phi(x, y) \leqq \phi(x', y')$ whenever $x \leqq x'$, $y \leqq y'$. It is easily seen that (ii) implies (i); we shall now show that the test ϕ that maximizes inf ω' $\beta(\theta_1, \theta_2)$ does not possess property (i) and hence also not (ii), provided $\theta_1^{**} - \theta_1^*$ and $\theta_2^{**} - \theta_2^*$ are sufficiently large so that ϕ is not identically equal to α. Let β denote the power function of ϕ and suppose that inf ω' $\beta(\theta_1, \theta_2) = \gamma > \alpha$. Then condition (i) implies that under the hypothesis $\beta(\theta_1, \theta_2) = \alpha$ only when $\theta_1 = \theta_1^*$ and $\theta_2 = \theta_2^*$. For if $\beta(\theta_1, \theta_2) = \alpha$ also for some other point in H, it would also equal α on the line segment connecting these two points and hence by analyticity on the whole line containing this segment. But this would imply that $\beta(\theta_1, \theta_2) = \alpha$ also for points in ω' where by assumption $\beta(\theta_1, \theta_2) \geqq \gamma > \alpha$. Another consequence of condition (i) is that $\beta(\theta_1, \theta_2) > \gamma$ for all points in ω' so that the minimum point γ is never attained in ω' and is approached only as either θ_1 or θ_2 tend to $-\infty$. For if, for example, $\beta(\theta_1', \theta_2') = \gamma$ for some point with $\theta_1' \geqq \theta_2^{**}$ and finite θ_1' we would have $\beta(\theta_1, \theta_2') = \gamma$ for all $\theta_1 \leqq \theta_1'$ and hence for all θ_1. This would imply $\beta(\theta_1, \theta_2) = \gamma$ for all (θ_1, θ_2) with $\theta_1 \geqq \theta_1^{**}$, $\theta_2 \leqq \theta_2'$ and therefore $\beta(\theta_1, \theta_2) \equiv \gamma$.

From these two remarks and Theorem 3.10 of Wald's book *Statistical Decision Functions* [3], it can be shown that there exists a sequence λ_i of probability distributions over ω' with the following properties: (a) For any real number A the probability under λ_i of the intersection of ω' with the quadrant $\{\theta_1, \theta_2 \mid \theta_1, \theta_2 \geqq A\}$ tends to zero as $i \to \infty$. (b) The power of the most powerful level α test for testing $H': \theta_1 = \theta_1^*$, $\theta_2 = \theta_2^*$ against the simple alternative $\int_{\omega'} p_{\theta_1,\theta_2}(x, y) d\,\lambda_i(\theta_1, \theta_2)$ tends to γ as $i \to \infty$. But from (a) it follows easily that as $i \to \infty$, $\int p_{\theta_1,\theta_2}(x, y) d\,\lambda_i(\theta_1, \theta_2)$ can be distinguished arbitrarily well from $p_{\theta_1^*,\theta_2^*}(x, y)$. This leads to the contradiction $\gamma = 1$.

We have given the proof explicitly only for the case of independent normal variables. However it applies equally well to any problem in which, in addition to the analyticity assumption of the present section, also the assumptions of Theorem 4.1 are satisfied.

4. Monotone regions. In the previous section we showed that for the hypotheses under consideration neither unbiasedness nor the minimax principle lead to desirable results. In order to arrive at a reasonable test we now impose the following preliminary conditions suggested by the negative results of the last section. We ask first that the test be nonrandomized, so that we can speak of a region w of rejection. The second restriction is one of monotonicity. Let us assume that we are concerned with two random variables X and Y whose joint distribution is given by $p_{\theta_1, \theta_2}(x, y)$. Then we shall say that the region w of rejection for the hypothesis $H: \theta_1 \leqq \theta_2^*, \theta_2 \leqq \theta_2^*$ is *monotone* (nondecreasing in x and y) if

$$(4.1) \qquad (x, y) \, \varepsilon \, w, \; x \leqq x', \; y \leqq y' \quad \text{imply} \quad (x', y') \, \varepsilon \, w,$$

that is, if its critical function is nondecreasing in both variables.

The restriction to monotone regions is of course suitable only in certain problems, namely, roughly speaking, when increased values of the parameters lead to higher values for the observations. To make this precise let $\theta_1 \leqq \theta_1', \theta_2 \leqq \theta_2'$ and let F and G be the cumulative distribution functions of X and Y corresponding to (θ_1, θ_2) and (θ_1', θ_2'), respectively. Then we shall consider the condition of monotonicity appropriate provided for every monotone non-decreasing region w

$$(4.2) \qquad \int_w dF \leqq \int_w dG.$$

Frequently the simplest way to prove (4.2) is to establish the existence of functions $x' = f(x, y)$, $y' = g(x, y)$ with $x' \geqq x$, $y' \geqq y$ and such that when F is the cumulative distribution function of (X, Y), that of (X', Y') is G. Sometimes it is more convenient instead to prove the existence of random variables, say $Z_1, \cdots, Z_r$, and functions $X = f(Z_1, \cdots, Z_r)$, $Y = g(Z_1, \cdots, Z_r)$, $X' = f'(Z_1, \cdots, Z_r)$, $Y' = g'(Z_1, \cdots, Z_r)$ such that $X \leqq X'$, $Y \leqq Y'$ and the cdf's of (X, Y) and (X', Y') are F and G respectively. Both of these conditions clearly assure the validity of (4.2) since for any w that is nondecreasing in x and y they imply

$$(4.3) \qquad \int_w dF = P((X, Y) \in w) \leqq P((X', Y') \in w) = \int_w dG.$$

A remark is required also in connection with the restriction to nonrandomized tests. When dealing with discrete problems, for example binomial distributions, we must permit a certain rather trivial kind of randomization. A formal way of handling the distinction is provided by a representation of randomized tests due to M. Eudey [4]. Let X denote the number of successes in n binomial trials, and let U be uniformly distributed over [0, 1]. Then any randomized test in X is equivalent to a non-randomized test in $X + U$, and we shall consider monotone non-randomized tests in the continualized variables $X + U$, $Y + V$. Here monotonicity insures that no very heavy use is made of randomization. In fact, in the original variables X, Y randomization will occur only on the boundary of the critical region.

We shall now derive that test of the hypothesis $H: \theta_1 \leqq \theta_1^*, \theta_2 \leqq \theta_2^*$ that among all monotone regions maximizes $\inf_{\omega'} \beta(\theta_1, \theta_2)$ where ω' is the set of points (θ_1, θ_2) with $\theta_1 \geqq \theta_1^{**}$ or $\theta_2 \geqq \theta_2^{**}$. Here $\theta_1^{**} \geqq \theta_1^*$ and $\theta_2^{**} \geqq \theta_2^*$. It may be of interest to point out that if we let $\theta_1^{**} = \theta_1^*$, $\theta_2^{**} = \theta_2^*$, we get the monotone region with minimum bias.

THEOREM 4.1. *Let the joint density of X and Y be* $p_{\theta_1,\theta_2}(x, y)$ *where the parameter-space is a finite or infinite open rectangle* $\underline{\theta}_1 < \theta_1 < \bar{\theta}_1$, $\underline{\theta}_2 < \theta_2 < \bar{\theta}_2$, *and the positive sample space also is an open rectangle* $\underline{x} < x < \bar{x}$, $\underline{y} < y < \bar{y}$ *independent of the θ's. Suppose that* (4.2) *holds, that the marginal distribution of X depends only on* θ_1, *and that of Y only on* θ_2, *and that X tends in probability to* $\underline{x}$ *as* $\theta_1 \to \underline{\theta}_1$, *while Y tends to* $\underline{y}$ *as* $\theta_2 \to \underline{\theta}_2$. *Then the test that among all monotone nonrandomized tests of H maximizes the minimum power against* ω' *is given by the region of acceptance S*:

$$(4.4) \qquad x \leqq a, \qquad y \leqq b,$$

where a and b are determined by the conditions

$$(4.5) \qquad P(X \leqq a, Y \leqq b \mid \theta_1^*, \theta_2^*) = 1 - \alpha$$

and

$$(4.6) \qquad P(X \leqq a \mid \theta_1^{**}) = P(Y \leqq b \mid \theta_2^{**}).$$

PROOF. We point out first that for any $x > \underline{x}$, $y > \underline{y}$

$$(4.7) \qquad \begin{aligned} \lim_{\theta_1 \to \underline{\theta}_1} P(X \leqq x, Y \leqq y \mid \theta_1, \theta_2) &= P(Y \leqq y \mid \theta_2) \\ \lim_{\theta_2 \to \underline{\theta}_2} P(X \leqq x, Y \leqq y \mid \theta_1, \theta_2) &= P(X \leqq x \mid \theta_1). \end{aligned}$$

For

$$P(X \leqq x, Y \leqq y) = P(X \leqq x) - P(X \leqq x, Y \geqq y)$$

while

$$0 \leqq P(X \leqq x, Y > y) \leqq P(Y > y)$$

and

$$\lim_{\theta_2 \to \underline{\theta}_2} P(Y > y \mid \theta_2) = 0.$$

For any monotone test the limit of the power $\beta(\theta_1, \theta_2)$ as $\theta_1 \to \underline{\theta}_1$ clearly exists and we shall denote it by $\beta(\underline{\theta}_1, \theta_2)$. The minimum power in ω' is then the smaller of the two quantities $\beta(\underline{\theta}_1, \theta_2^{**})$ and $\beta(\theta_1^{**}, \underline{\theta}_2)$. Since for the test given by (4.5) we have $\beta(\underline{\theta}_1, \theta_2^{**}) = \beta(\theta_1^{**}, \underline{\theta}_2)$ we could, if the theorem were false, increase both $\beta(\underline{\theta}_1, \theta_2^{**})$ and $\beta(\theta_1^{**}, \underline{\theta}_2)$.

Clearly any monotone test has a region of acceptance S' of the form $y \leqq g(x)$

or equivalently $x \leqq h(y)$ where g and h are non-increasing. If $S' \neq S$ is to be of size $\leqq \alpha$ we have $P(S' \mid \theta_1^*, \theta_2^*) \geqq P(S \mid \theta_1^*, \theta_2^*)$ and hence either $\lim_{x \to \bar{x}} g(x) > b$ or $\lim_{y \to \bar{y}} h(y) > a$. Suppose that the first of these conditions is satisfied and let us denote the complement of S and S' by $\bar{S}$ and $\bar{S}'$, respectively. Then there exists a constant k such that $x \geqq k$ for all points in $S \wedge \bar{S}'$, and a subset S'' of $\bar{S} \wedge S'$ such that $P(S'' \mid \theta_1, \theta_2) > 0$ for all θ_1, θ_2 and $x \leqq k$ for all points in S″. Hence by (4.7)

$$P(S \wedge \bar{S}' \mid \underline{\theta}_1, \theta_2^{**}) = 0$$

and

$$P(\bar{S} \wedge S' \mid \underline{\theta}_1, \theta_2^{**}) = P(b \leqq Y \leqq \lim_{x \to \bar{x}} g(x) \mid \theta_2^{**}) > 0,$$

which leads to

$$P(S' \mid \underline{\theta}_1, \theta_2^{**}) > P(S \mid \underline{\theta}_1, \theta_2^{**}),$$

and hence to the desired result.

The theorem becomes particularly simple if the joint density of X and Y is symmetric in its two variables when $\theta_1 = \theta_2$. For then if $\theta_1^* = \theta_2^* = \theta^*$, $\theta_1^{**} = \theta_2^{**} = \theta^{**}$, it is seen from (4.6) that $a = b$. Thus the test accepts if $\max(X, Y) \leqq a$ where a is determined by (4.5), and hence is independent of θ^{**}.

The assumptions made in Theorem 4.1 concerning the shape of the parameter and sample spaces are unnecessarily restrictive. The theorem remains valid if we assume that both the parameter space and the positive sample space are convex open sets. The proof is essentially the same, however the notation becomes considerably more complicated.

If the roles of hypothesis and class of alternatives are interchanged, we obtain

THEOREM 4.2. *For testing the hypothesis $H': \theta_1 \geqq \theta_1^{**}$ or $\theta_2 \geqq \theta_2^{**}$ against the class of alternatives $\omega: \theta_1 \leqq \theta_1^*$, $\theta_2 \leqq \theta_2^*$, let S be the region of rejection $x \leqq c$, $y \leqq d$, where c and d are determined by*

$$P(X \leqq c \mid \theta_1^{**}) = P(Y \leqq d \mid \theta_2^{**}) = \alpha. \tag{4.8}$$

Then under the assumptions of Theorem 4.1 S is uniformly most powerful among all regions of rejection that are monotone non-increasing in both variables.

PROOF. Consider any monotone region given by $x \leqq g(y)$ or $y \leqq h(x)$ with g and h non-increasing. Since the probability of rejection must be not greater than α at $(\theta_1^{**}, \underline{\theta}_2)$ and $(\underline{\theta}_1, \theta_2^{**})$ we must have

$$\lim_{x \to \bar{x}} g(x) \leqq c, \qquad \lim_{y \to \bar{y}} h(y) \leqq d. \tag{4.9}$$

But any monotone region satisfying (4.9) is contained in S and hence is uniformly less powerful than S.

5. Examples. In the present section we shall apply Theorem 4.1 to some specific problems. All other assumptions being trivially satisfied we shall in each case only check condition (4.2).

Example 5.1. Let X, Y have a multinomial distribution

$$(4.10) \quad P(X = x, Y = y) = \frac{n!}{x!y!(n - x - y)!} \theta_1^x \theta_2^y (1 - \theta_1 - \theta_2)^{n-x-y}.$$

To see that (4.2) holds let $Z_1, \cdots, Z_n$ be uniformly distributed on (0, 1) and let U, V denote the number of Z's in the intervals $(0, \theta_1)$ and $(1 - \theta_2, 1)$ respectively. If U', V' are defined analogously with respect to θ_1', θ_2' it is seen that (U, V) has the distribution (4.10) while (U', V') has the corresponding distribution for θ_1', θ_2'. Since $U \leqq U'$, $V \leqq V'$ the validity of (4.2) follows.

The same proof works of course in the case that X and Y are independently distributed, each according to a binomial distribution.

Example 5.2. Let $X_1, \cdots, X_n$ be independently and normally distributed with mean ξ and variance σ^2, and consider the hypothesis $H: \xi \leqq \xi_0$, $\sigma \leqq \sigma_0$. Since $\bar{X}$ and $S^2 = \sum(X_i - \bar{X})^2$ are sufficient for ξ and σ, we may restrict attention to these statistics. However, if we try to set $X = \bar{X}$, $Y = S^2$, $\theta_1 = \xi$, $\theta_2 = \sigma^2$ we encounter certain difficulties. First the distribution of X does not depend only on θ_1 as we require in Theorem 4.1. While this is not a very important condition of the theorem, a second consideration shows that it is impossible to apply the monotonicity restriction at all to the present set-up. For the joint cumulative distribution function of $\bar{X}$ and S does not satisfy condition (4.2).

This exhibits an unpleasant feature of the present approach. In a given problem it is not known a priori whether there will exist variables X, Y and a choice of the parameters θ_1, θ_2 so that (4.2) will be satisfied. On the other hand, when such variables and parameters have been found, it is not clear that these are the only possible choices. While it would of course be interesting to investigate existence and uniqueness questions, the monotonicity condition is an extraneous restriction anyway, whose suitability must be judged for each problem in terms of the choices for X, Y, θ_1, and θ_2.

In the present case we may take $X = (\bar{X} - \xi_0)/S$, $Y = S^2$, $\theta_1 = (\xi - \xi_0)/\sigma$ and $\theta_2 = \sigma^2$. To check condition (4.2) assume without loss of generality that $\xi_0 = 0$ and let $\theta_1 = \xi/\sigma < \xi'/\sigma' = \theta_1'$, $\sigma < \sigma'$. If $\sigma' = k\,\xi' = k\xi + c$, let $X_i' = kX_i + c$, so that $S' = kS$, $\bar{X}' = k\bar{X} + c$. Since $k > 1$ and $\xi < \xi'/k$ we see that $c > 0$ and $\bar{X}' > \bar{X}$, $Y' > Y$, so that (4.2) follows.

As a last problem let us consider one in which nuisance parameters are present.

Example 5.3. Let $X_1, \cdots, X_m$; $Y_1, \cdots, Y_n$ be independently normally distributed with common variance σ^2; let $E(X_i) = \xi$, $E(Y_j) = \eta$, and consider the hypothesis $H: \xi \leqq \xi_0$, $\eta \leqq \eta_0$. This time $\bar{X}$, $\bar{Y}$, and $S^2 = \sum(X_i - \bar{X})^2 + \sum(Y_j - \bar{Y})^2$ from a set of sufficient statistics, and again the question arises how to choose X, Y, θ_1, and θ_2. Here the principle of invariance (see [5]) leads to a solution very simply. If one rewrites the hypothesis in the form: $(\xi - \xi_0)/\sigma \leqq 0$, $(\eta - \eta_0)/\sigma \leqq 0$ it is seen that $X = (\bar{X} - \xi_0)/S$, $Y = (\bar{Y} - \eta_0)/S$ constitute a maximal invariant under a suitable group of transformations. The corresponding parameter invariants are of course $\theta_1 = (\xi - \xi_0)/\sigma$, $\theta_2 = (\eta - \eta_0)/\sigma$.

It remains, once more, to check (4.2). For this purpose let ξ, η, ξ', η', σ be numbers such that

$$\frac{\xi}{\sigma} = \theta_1, \quad \frac{\eta}{\sigma} = \theta_2, \quad \frac{\xi'}{\sigma} = \theta_1', \quad \frac{\eta'}{\sigma} = \theta_2'$$

and let $U_1, \cdots, U_m; V_1, \cdots, V_n$ be independently normally distributed with common variance σ^2 and means $E(U_i) = \xi$, $E(V_j) = \eta$. If $T^2 = \sum(U_i - \bar{U})^2 + \sum(V_j - \bar{V})^2$, the joint distribution of $U = \bar{U}/T$, $V = \bar{V}/T$ is that corresponding to (θ_1, θ_2). Let $\xi' - \xi = c$, $\eta' - \eta = d$ and let $U_i' = U_i + c$, $V_j' = V_j + d$. Then $U' = \bar{U}'/T$, $V' = \bar{V}'/T$ have the joint distribution corresponding to (θ_1', θ_2'). Also $U' = \bar{U}+c/T > U$ and $V' = \bar{V}+c/T > V$.

Exactly analogously we can treat the hypothesis $H'{:}\xi/\sigma \leqq \gamma$, $\eta/\sigma \leqq \delta$, and the corresponding problems in which the two variances are not assumed to be equal.

6. A multidecision problem. As was pointed out in the introduction, some of the problems considered here really involve the choice between more than two decisions. We shall now indicate, by discussing an example, one method of treating such multidecision problems through successive reduction to problems of the simpler type.

Let us once more consider the hypothesis H : $\theta_1, \theta_2 \leq 0$ and let us assume that in case of rejection we wish to decide whether $\theta_2 > 0 \geq \theta_1$, $\theta_1 > 0 \geq \theta_2$ or whether θ_1 and θ_2 are both > 0[2]. Let us denote these three regions of the parameter space by ω_1, ω_2, and ω_3 and the associated decisions by d_1, d_2, d_3. The set $\theta_1, \theta_2 \leq 0$ will be denoted by ω_0 and the associated decision of accepting H by d_0.

We shall assume that each of the four pairs of random variables $(\pm X, \pm Y)$ is monotone with respect to the corresponding pair of parameters $(\pm\theta_1, \pm\theta_2)$ and that $p_{\theta,\theta}(x, y)$ is symmetric in x and y so that the region of acceptance is given by

$$\max(x, y) \leqq a. \tag{6.1}$$

We must now consider how to divide up the complementary region between d_1, d_2, and d_3. Here we again impose the natural monotonicity restrictions. We ask that the region for d_1 be monotone non-increasing in x and non-decreasing in y, and that the analogous conditions be satisfied by the regions for d_2 and d_3. Suppose the problem concerns a standard treatment and two new ones, where θ_1 and θ_2 measure in some way the differences between the new treatments and the standard. The circumstances are such that the most serious error consists in incorrectly rejecting the standard treatment in favor of one of the others. By proper choice of a this probability is controlled so that it is not greater than α for all $(\theta_1, \theta_2) \, \varepsilon \, \omega_0$.

Next in importance seems to come the possibility of reaching decision d_2 in

[2] A similar multidecision problem has recently been treated by Paulson [6] from a somewhat different point of view.

ω_1 or d_1 in ω_2, but we shall set these aside for the moment. The next important error presumably consists in deciding on d_3 when the parameter point lies either in ω_1 or ω_2. This error can be controlled in the usual manner by making the d_3-region sufficiently small. Thus we may select a number $0 < \beta < 1$ and impose bounds

$$P(d_3 \mid \omega_1) \leqq \beta, \qquad P(d_3 \mid \omega_2) \leqq \beta. \tag{6.2}$$

Subject to these conditions we wish to maximize $P(d_3)$ in ω_3. Let us now restrict attention to d_3-regions of the type $y \geqq g(x)$ or $x \geqq h(y)$ with g and h non-increasing. Then it is seen by the argument used to prove Theorem 4.2 that among all monotone d_3-regions satisfying (6.2) there exists one that uniformly maximizes $P(d_3)$ over ω_3. If $P(X > a \mid \theta_1 = 0) \leqq \beta$ it consists of the points satisfying either $x \geqq a$, $y \geqq b$ or $x \geqq b$, $y \geqq a$ where a, b are determined by

$$P(X \leqq a, Y \leqq a \mid \theta_1 = \theta_2 = 0) = 1 - \alpha,$$

$$P(X > b \mid \theta_1 = 0) = \beta.$$

If on the other hand $P(X > a \mid \theta_1 = 0) \geqq \beta$ the optimum d_3-region is given by $x \geqq b$, $y \geqq b$.

Let the remainder of the sample space be divided up symmetrically in the obvious manner between d_1 and d_2. It then follows from monotonicity that $P(d_1 \mid \omega_2)$ and $P(d_2 \mid \omega_1)$ both take on their maximum value at $\theta_1 = \theta_2 = 0$. Hence $P(d_1 \mid \omega_2) \leqq \frac{1}{2}\alpha$, $P(d_2 \mid \omega_1) \leqq \frac{1}{2}\alpha$, so that these errors also are controlled in a satisfactory manner.

7. Convex regions.[3] If we try to apply the results of Section 4 to specific examples, we occasionally find an obstacle in the condition that X and Y should tend in probability to $\underline{x}$ and $\underline{y}$ as $\theta_1 \to \underline{\theta}_1$ and $\theta_2 \to \underline{\theta}_2$, respectively. We shall now show that by restricting the acceptance region to be convex as well as monotone we can prove a result analogous to Theorem 4.1 without assuming degeneracy of the distribution at θ_1 and θ_2.

Let us consider again the joint density $p_{\theta_1,\theta_2}(x, y)$ satisfying (4.2), the hypothesis $H: \theta_1 \leqq \theta_1^*$, $\theta_2 \leqq \theta_2^*$ and the set of alternatives $\theta_1 \geqq \theta_1^{**}$ or $\theta_2 \geqq \theta_2^{**}$. Putting

$$r_\pi(x, y) = \frac{p_{\theta_1^*,\theta_2^*}(x, y)}{\pi p_{\theta_1,\theta_2^{**}}(x, y) + (1 - \pi) p_{\theta_1^{**},\theta_2}(x, y)},$$

we shall assume:

(i) For any $0 < \pi < 1$ and any C the region

$$r_\pi(x, y) \geqq C \tag{7.1}$$

is convex and non-decreasing in x and y.

[3] I am indebted to the referee for several valuable suggestions with regard to this section.

(ii) Given any two points (x', y') and (x'', y'') with $x' < x''$, $y' > y''$ there exists $0 < \pi < 1$ and C such that both points lie on the boundary of the set of points (x, y) satisfying (7.1).

(iii) $p_{\theta_1, \theta_2}(x, y) > 0$ for all θ_1, θ_2, x, y under consideration.

Consider now the set of alternatives $\omega': \theta_1 \geqq \underline{\theta}_1$, $\theta_2 \geqq \theta_2^{**}$ or $\theta_2 \geqq \underline{\theta}_2$, $\theta_1 \geqq \theta_1^{**}$ where θ_1 and θ_2 may now be any numbers less than θ_1^* and θ_2^*, respectively. Let a and b be determined by

$$P(X \leqq a, Y \leqq b \mid \theta_1^*, \theta_2^*) = 1 - \alpha,$$

$$P(X \leqq a, Y \leqq b \mid \underline{\theta}_1, \theta_2^{**}) = P(X \leqq a, Y \leqq b \mid \theta_1^{**}, \underline{\theta}_2).$$

Then the acceptance region S: $x \leqq a, y \leqq b$ maximizes $\inf_{\omega'} \beta(\theta_1, \theta_2)$ among all monotone and convex level α tests.

For let S' be any other acceptance region satisfying the conditions that have been imposed. The boundary curves of S and S' have in common either one or

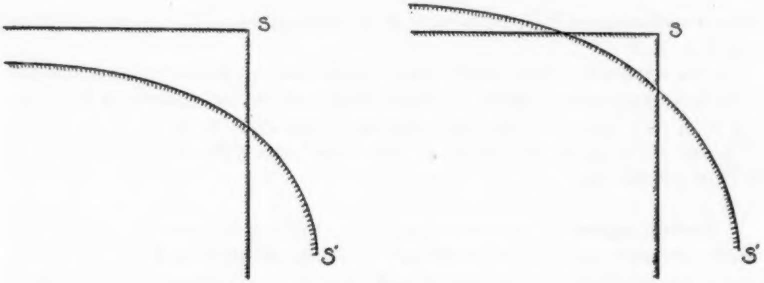

FIG. 1.

two points or an interval. Let us consider the case of two points, say (x', y') and (x'', y''). We may then assume $x' < x''$, $y' > y''$ so that there exist π and C such that the boundary of (7.1) passes through these two points. From (i) it follows that $r(x, y) \geqq C$ for all points in $S \wedge \bar{S}'$ and $\leqq C$ in $\bar{S} \wedge S'$. Since we have $P(S') \leqq P(S) = \alpha$ when the density of X, Y is $p_{\theta_1^*, \theta_2^*}(x, y)$ it follows from the fundamental lemma of Neyman and Pearson that $P(S') > P(S)$ when the density is given by $\pi p_{\theta_1^*, \theta_2^{**}}(x, y) + (1 - \pi) p_{\theta_1^{**}, \theta_2}(x, y)$. It is therefore impossible that $P(S') \geqq P(S)$ for both $(\underline{\theta}_1, \theta_2^{**})$ and $(\theta_2^{**}, \underline{\theta}_2)$ as was to be proved.

The same argument applies also in the case that the boundaries of S and S' have an interval in common. Consider finally the case of one common point, say (a, y_0). For each n let (π_n, C_n) satisfy (ii) for $(-n, b)$ and (a, y_0). Then

$$0 \leqq C_n \leqq \max \left\{ \frac{p_{\theta_1^*, \theta_2^*}(a, y_0)}{p_{\underline{\theta}_1, \theta_2^{**}}(a, y_0)}, \frac{p_{\theta_1^*, \theta_2^*}(a, y_0)}{p_{\theta_1^{**}, \underline{\theta}_2}(a, y_0)} \right\},$$

so that there is a subsequence of $\{(\pi_n, C_n)\}$ which converges, say, $\pi_n \to \pi^*$, $C_n \to$

C^*. It is easily seen from the monotonicity and convexity of the regions $r_{\pi_n}(x, y) \geqq C_n$ that the boundary of $r_{\pi^*}(x, y) \geqq C^*$ is the line $y = y_0$. The remainder of the argument is completely analogous to the two point case.

As an example let

$$p_{\theta_1, \theta_2}(x, y) = C(\theta_1, \theta_2)\, e^{\theta_1 x + \theta_2 y}\, h(x, y), \tag{7.2}$$

and assume that (4.2) holds. Suppose that a priori lower bounds $\underline{\theta}_1$, $\underline{\theta}_2$ are given for θ_1, θ_2 such that $p_{\underline{\theta}_1, \underline{\theta}_2}(x, y)$ is again a density. If we let $\theta_1^{**} = \theta_1^*$, $\theta^{**} = \theta_2^*$, the region (7.1) becomes

$$a\, e^{(\theta_1 - \dot{\theta}_1)x} + b\, e^{(\theta_2 - \dot{\theta}_2)y} \leqq k$$

and conditions (i), (ii), and (iii) are easily checked.

8. Two-sided problems. We shall discuss only one rather trivial two-sided problem, which is enough however to indicate that the type of result one obtains here is entirely different from what we found in the one-sided case.

Let X and Y be independently, normally distributed with unit variance and means ξ and η, respectively, and consider the hypothesis $H: \xi = \eta = 0$. We shall determine the test ϕ that maximizes $\inf_{\omega'} \beta(\xi, \eta)$ where ω' is the set of points for which either $|\xi|$ or $|\eta|$ is $\geqq \gamma$, $(\gamma > 0)$. Any reasonable test for this problem would presumably attain its minimum power over ω' at the four points $(0, \gamma)$, $(0, -\gamma)$, $(\gamma, 0)$, $(-\gamma, 0)$. We therefore expect ϕ to be the most powerful test of H against the simple alternative that assigns probability $\frac{1}{4}$ to each of these 4 points. The region of acceptance for this problem is given by S:

$$e^{\gamma x} + e^{-\gamma x} + e^{\gamma y} + e^{-\gamma y} \leqq k.$$

It is easily checked that this has the following properties:

(i) S is convex.

(ii) If $0 \leqq x' \leqq x$, $0 \leqq y' \leqq y$ and $(x, y) \,\varepsilon\, S$, the point (x', y') also lies in S.

(iii) For any fixed η_0 the probability of S decreases with $|\xi|$ and for fixed ξ_0 decreases in $|\eta|$.

From (iii) it follows that ϕ is the test we are looking for, and it seems to be entirely satisfactory. In fact, if we utilize the symmetry of the situation to reduce the variables to $|X|$, $|Y|$ and the parameters to $|\xi|$, $|\eta|$ we are faced essentially with a one-sided situation and it is seen from (i) and (ii) that the acceptance region, when interpreted in this way, is both monotone and convex.

9. A general concept of monotonicity. In Sections 4–6 we made use of the notion of monotonicity, and we shall conclude this paper by indicating how this concept may be extended to the general decision problem.

Suppose that there is defined a partial ordering $\leqq$ in the sample space and a partial ordering $\dot{\leqq}$ in the parameter-space. In analogy to condition (4.2) we shall assume that if $\mathcal{W}$ is any monotone non-decreasing region in the sample space we have for any two parameter-points $\theta \dot{\leqq} \theta' : P_\theta(\mathcal{W}) \leqq P_{\theta'}(\mathcal{W})$.

Suppose that there is also defined a partial ordering in the decision space, to be denoted by $\underset{\cdot}{\leqq}$. We shall assume that the loss function W satisfies the conditions:

(i) $d_1 \underset{\cdot}{\leqq} d_2 \underset{\cdot}{\leqq} d_3$ and $W(\theta, d_1) < W(\theta, d_2)$ implies $W(\theta, d_2) \leqq W(\theta, d_3)$,

(ii) $\theta_1 \dot{\leqq} \theta_2 \dot{\leqq} \theta_3$ and $W(\theta_1, d) < W(\theta_2, d)$ implies $W(\theta_2, d) \leqq W(\theta_3, d)$. Under these assumptions it seems natural to restrict consideration to monotone decision functions, where we shall call δ monotone if $x \leqq x'$ implies $\delta(x) \underset{\cdot}{\leqq} \delta(x')$.

REFERENCES

[1] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses. Part III." *Statistical Research Memoirs*, Vol. 2 (1938), pp. 36–57.

[2] Stanley L. Isaacson, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 217–234.

[3] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

[4] Mark W. Eudey, "On the treatment of discontinuous random variables," *Technical Report No. 13, ONR Project NR-042-036*, 1949.

[5] E. L. Lehmann, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 1–26.

[6] Edward Paulson, "On the comparison of several experimental categories with a control," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 239–246.

IMPARTIAL DECISION RULES AND SUFFICIENT STATISTICS[1]

By Raghu Raj Bahadur and Leo A. Goodman

The University of Chicago

Summary. A class of decision problems concerning k populations was considered in [1] and it was shown that a particular decision rule is the uniformly best 'impartial' decision rule for many problems of this class. The present paper provides certain improvements of this result. The authors define impartiality in terms of permutations of the k samples rather than in terms of the k ordered values of an arbitrarily chosen real-valued statistic as in the earlier paper. They point out that (under conditions which are satisfied in the standard cases of k independent samples of equal size) if the same function is a sufficient statistic for each of the k samples then the conditional expectation of an impartial decision rule given the k sufficient statistics is also an impartial decision rule. A characterization of impartial decision rules is given which relates the present definition of impartiality with the one adopted in [1]. These results, together with Theorem 1 of [1], yield the desired improvements. The argument indicated here is illustrated by application to a special case.

1. Introduction. Let $\pi_1, \pi_2, \cdots, \pi_k$ be a given set of populations and let the distribution function of a single observation x_i from π_i be

$$\Pr(x_i \leqq z) = G(z, \theta_i) \qquad (-\infty < z < \infty), \tag{1}$$

where θ_i is an unknown parameter (not necessarily real-valued), $i = 1, 2, \cdots, k$. Write $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$ and let Ω be the set of all points ω which are regarded as being possible in the given case. Suppose that n independent observations are drawn from π_i, giving the sample $(x_{i1}, x_{i2}, \cdots, x_{in}) = u_i$ say, $i = 1, 2, \cdots, k$, and let the combined sample point $(u_1, u_2, \cdots, u_k)$ be denoted by v. Let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be an ordered set of functions p_i of the combined sample point v such that

$$0 \leqq p_i(v) \leqq 1, \qquad \sum_{i=1}^{k} p_i(v) = 1 \tag{2}$$

for all v. Then $d(v)$ is said to be a *decision rule*. The statistical problems which motivate this definition may be described as follows.

Suppose that it is desired to determine appropriate sampling rates $p_1, p_2, \cdots, p_k$ for $\pi_1, \pi_2, \cdots, \pi_k$, respectively, p_i being the relative proportion of x's which will be drawn in future from π_i, $(0 \leqq p_i \leqq 1, \sum_1^k p_i = 1)$. For example, the given populations may be k varieties of grain, x_i the yield (bushels per acre or dollars profit per acre) from π_i, and p_i the proportion of the available land on which π_i is to be grown, $i = 1, 2, \cdots, k$. Again, $\pi_1, \pi_2, \cdots, \pi_k$ may be

[1] This paper was prepared in connection with research sponsored by the Office of Naval Research.

sources of a manufactured article, x_i the relevant quality characteristic (e.g., number of hours of service) of an article supplied by π_i, and $p_1, p_2, \cdots, p_k$ the relative proportions in which a consumer obtains the articles he needs from $\pi_1, \pi_2, \cdots, \pi_k$, respectively. The mixed population obtained by using a given set $d^0 = (p_1^0, p_2^0, \cdots, p_k^0)$ of sampling rates is characterized by the distribution function

$$G(z \mid \omega, d^0) = \sum_{i=1}^{k} p_i^0 G(z, \theta_i) \qquad (-\infty < z < \infty), \tag{3}$$

where the component distribution functions $G(z, \theta_i)$ are given by (1), and the object is to determine d^0 in such a way that $G(z \mid \omega, d^0)$ has properties which are desirable in the given case. (For instance, it may be desirable to minimize $G(a \mid \omega, d^0)$, where a is a given constant, or to maximize $G(b + \epsilon \mid \omega, d^0) - G(b - \epsilon \mid \omega, d^0)$ where b and $\epsilon > 0$ are given constants.) If the parameters θ_i were known, an appropriate d^0 could (presumably) be determined, but otherwise the statistician must resort to sampling the populations and take d^0 to be a function of the sample values. If samples of fixed size n are drawn from each population, we see that the statistician will be using a decision rule, say $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$. The expected distribution function of the mixed population obtained by using the rule $d(v)$ is (cf. (3))

$$\begin{aligned} H(z \mid \omega, d) &= E[G(z \mid \omega, d(v)) \mid \omega] \\ &= \sum_{i=1}^{k} G(z, \theta_i) E[p_i(v) \mid \omega] \qquad (-\infty < z < \infty), \end{aligned} \tag{4}$$

where $E[p_i(v) \mid \omega]$ denotes the expected value of $p_i(v)$ when the true parameter point is ω. The statistician's problem is to construct a decision rule $d^*(v)$ such that $H(z \mid \omega, d^*)$ has properties which are desirable in the given case.

A special version of applications of this type is the following. For brevity, write $G_i(z) = G(z, \theta_i)$ and let $\lambda(G)$ be a real-valued functional on the distribution functions $G_i(z)$, for example, $\lambda(G) = \int_{-\infty}^{\infty} z \, dG(z)$ or $\lambda(G) = G(b + \epsilon) - G(b - \epsilon)$, where b and $\epsilon > 0$ are given constants. Writing $\lambda_i = \lambda(G_i)$, suppose that it is desired to select a population, π_r say, from the given set $\pi_1, \pi_2, \cdots, \pi_k$ such that $\lambda_r = \max\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. Since the G_i's are unknown, it will in general be impossible to effect a (or the) correct selection with certainty, but if it is agreed to make the selection depend on the outcome of drawing samples of size n from each population, the most general selection procedure is to use a decision rule, $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ say, in the following manner: given v, the statistician performs a random experiment whose outcome ρ takes on only the values $1, 2, \cdots, k$ with

$$\Pr(\rho = i) = p_i(v) \qquad (i = 1, 2, \cdots, k)$$

and selects π_ρ. The probabilities of selecting $\pi_1, \pi_2, \cdots, \pi_k$ are then

$$E[p_1(v) \mid \omega], E[p_2(v) \mid \omega], \cdots, E[p_k(v) \mid \omega], \tag{5}$$

respectively. The problem in such a case might be to construct a decision rule $d^*(v)$ such that the probability of a correct selection in using $d^*(v)$ is "as large as possible."

In view of the applications (cf. (4), (5)) we shall say that two decision rules $d(v) = (p_1(v), \cdots, p_k(v))$ and $d'(v) = (p_1'(v), \cdots, p_k'(v))$ are *equivalent* if $E[p_i(v) \mid \omega] = E[p_i'(v) \mid \omega]$ for $i = 1, 2, \cdots, k$ and all ω in Ω.

We shall concern ourselves primarily with a class of decision rules which seems to be of interest on intuitive grounds. This is the class of impartial decision rules (see [1], [2]). Let us consider the case $k = 2$. Then a decision rule $d(v)$ is said to be impartial if $d(u_1, u_2) = (\alpha, \beta)$ implies $d(u_2, u_1) = (\beta, \alpha)$. In other words, $d(v) = (p_1(v), p_2(v))$ is an impartial decision rule if $p_1(u_2, u_1) = p_2(u_1, u_2)$, $p_2(u_2, u_1) = p_1(u_1, u_2)$ for all $v = (u_1, u_2)$. In the general case, a decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ is said to be *impartial* if for any $v = (u_1, u_2, \cdots, u_k)$ and any permutation $i_1 i_2 \cdots i_k$ of $1\ 2\ 3 \cdots k$ we have $p_j(u_{i_1}, u_{i_2}, \cdots, u_{i_k}) = p_{i_j}(u_1, u_2, \cdots, u_k)$ for $j = 1, 2, \cdots, k$.

The main result of this paper (Theorem 2) applies to cases whose essential feature is the existence of a function $s(u)$, not necessarily real-valued, on n dimensional sample space such that $s_i = s(u_i)$ is a sufficient statistic for θ_i $(i = 1, 2, \cdots, k)$, and such that the conditional distribution of u_i with s_i fixed equal to c is the same for each i. (The necessary conditions are always satisfied if, for example, the k populations are all (i) normal, or (ii) rectangular, or (iii) exponential, or (iv) binomial, or (v) of Poisson type.) Then $t(v) = (s_1, s_2, \cdots, s_k)$ is a sufficient statistic for ω, and it is clear, upon taking conditional expectations, that corresponding to any decision rule $d(v)$ there exists a decision rule $d^*(t(v))$ which is equivalent to $d(v)$ (Theorem 1). It is not immediately obvious, however, that if $d(v)$ is an impartial decision rule then this equivalent rule $d^*(t(v))$ will also be impartial. We show, in proving Theorem 2, that this is indeed the case. The question raised on page 374 of [1] with reference to Example 2 of that paper is thus answered in the affirmative. Our final result (Theorem 3) gives a characterization of impartial decision rules which relates the present definition to the one adopted in [1]. It might be pointed out that impartiality is a special case of invariance (cf. [6]), so that this is a special case of the following proposition: the conditional expectation of an invariant decision rule is also an invariant decision rule. A discussion of the general proposition will appear elsewhere.

Now, it is known (Theorem 1 of [1]) that there exist two impartial decision rules, called $d_1^*(t(v))$ and $d_k^*(t(v))$, such that in many applications $d_1^*(t(v))$ is the worst one and $d_k^*(t(v))$ the best one in the class of all impartial decision rules of the form $d^*(t(v))$ whatever the unknown parameter point ω may be; that is, $d_1^*(t(v))$ and $d_k^*(t(v))$ are the uniformly worst and uniformly best decision rules in the class of all impartial decision rules which are based on the sufficient statistics $s_1, s_2, \cdots, s_k$ alone. Theorem 2 shows that in these applications $d_1^*(t(v))$ and $d_k^*(t(v))$ are in fact uniformly worst and uniformly best in the class of *all* impartial decision rules. (Theorem 1 of [1] is stated and proved only in the "continuous" case, but can be extended to cover the discrete case as well; the necessary modifications become evident upon comparing Theorem 3 of the present paper with the development in [1].) By way of illustration of the argument indicated here, in the final section of the paper we consider certain problems connected with the case when $\pi_1, \pi_2, \cdots, \pi_k$ are normal populations having unknown means m_i and a common variance σ^2 (which may or may not be known) and prove a result (Theorem 4) which generalizes Example 1 of [1] as also a result due to Simon [2] for the case $k = 2$.

2. Theorems. The reader is referred to [3] for an account of such measure-theoretic terms and results as we use without explanation in what follows. Throughout this section we write '(sub)set' for 'Borel-measurable (sub)set' and 'function' for 'Borel-measurable function.' Functions whose range is not specified are understood to be real-valued. R^q denotes a *fixed subset* of the set of all points $z = (x_1, x_2, \cdots, x_q)$ with real coordinates x_i. (In our discussion, some of the spaces R^q will be given at the outset, and all the others will be defined explicitly in terms of them.) For any subset A of R^q, $\chi_A(z)$ denotes the characteristic function of A; that is, $\chi_A(z) = 1$ for z in A and $= 0$ for z in $R^q - A$. Let f be a nonnegative function on R^q and let λ be a measure on R^q (more precisely, a measure on the subsets of R^q) such that

$$\int_{R^q} f(z)\, d\lambda = 1.$$

Let Z be a random variable taking values in R^q such that the probability of event $\{Z \varepsilon A\}$ is

$$\int_A f(z)\, d\lambda$$

for all sets A. We then say that Z is distributed (on R^q) according to $f(z)\, d\lambda$.

Let $U_1, U_2, \cdots$, and U_k be independent random variables whose joint distribution is governed by a parameter ω taking values in a space Ω. Each U_i takes values in a set R^n of points u. Let s be a function on R^n onto a set R^m of points y. Let $h(u)$ be a nonnegative function of u, and let μ be a σ-finite measure on R^n. Corresponding to each ω in Ω and each $i = 1, 2, \cdots, k$ let $g_i(y:\omega)$ be a nonnegative function of y such that

$$\int_{R^n} h(u) g_i(s(u):\,\omega)\, d\mu = 1.$$

It is assumed that U_i is distributed according to $h(u)g_i(s(u):\omega)\, d\mu$ $(i = 1, 2, \cdots k)$.

Let R^{nk} $(= R^n \times R^n \times \cdots \times R^n)$ be the set of all points $v = (u_1, u_2, \cdots, u_i)$ with u_i in R^n $(i = 1, 2, \cdots, k)$, and write

$$\alpha(v) = \prod_{i=1}^{k} h(u_i),$$

$$t(v) = (s(u_1), s(u_2), \cdots, s(u_k)),$$

$$\beta(t(v):\,\omega) = \prod_{1=1}^{k} g_i(s(u_i):\,\omega).$$

Let $\mu^{(k)}$ be the product measure $\mu \times \mu \times \cdots \times \mu$ on R^{nk}. Then $V = (U_1, U_2, \cdots, U_k)$ is distributed according to $\alpha(v)\ \beta(t(v):\omega)\ d\mu^{(k)}$. If ϕ is a function on R^{nk}, we shall denote the expected value (if it exists) of $\phi(V)$, that is, the integral

$$\int_{R^{nk}} \phi(v)\alpha(v)\beta(t(v):\omega)\ d\mu^{(k)},$$

by $E[\phi(v) \mid \omega]$.

Let R^{mk} $(= R^m \times R^m \times \cdots \times R^m)$ be the set of all values of t as v ranges over R^{nk}, and let the generic point of R^{mk} be denoted by w or by $(y_1, y_2, \cdots, y_k)$. It can be shown that the preceding assumptions and definitions imply that t is a sufficient statistic for ω when the sample space is R^{nk}; that is, corresponding to each subset A of R^{nk} there exists a function ϕ_A, $0 \leqq \phi_A \leqq 1$, on R^{mk} such that

$$E[\chi_B(t(v))\ \chi_A(v) \mid \omega] = E[\chi_B(t(v))\ \phi_A(t(v)) \mid \omega]$$

for all subsets B of R^{mk} and all ω in Ω. ($\phi_A(w)$ is called the conditional probability of the event $\{V \ \varepsilon\ A\}$ given $t(V) = w$ and any ω in Ω). This property of t does not, however, suffice for our purpose; we require in addition the following result concerning the structure of the functions $\phi_A(w)$.

LEMMA. *Corresponding to each y in R^m there exists a probability measure λ_y on R^n such that for each A and w we may take $\phi_A(w) = \nu_w(A)$, where, for fixed $w = (y_1, y_2, \cdots, y_k)$, ν_w is the product measure $\lambda_{y_1} \times \lambda_{y_2} \times \cdots \times \lambda_{y_k}$ on R^{nk}.*

A proof of the lemma can be constructed along the following lines. (i) There exist functions $g_i(s:\omega)$ and a fixed probability measure μ^* on R^n such that U_i is distributed according to $g_i(s(u):\omega)\ d\mu^*$ $(i = 1, 2, \cdots, k; \omega\ \varepsilon\ \Omega)$. (ii) There exist functions $\lambda_y(C)$ such that, for each y, λ_y is a probability measure on R^n, and for each subset C of R^n, $\lambda_y(C)$ is the conditional measure of C given $s(u) = y$ when μ^* is the unconditional measure on R^n (cf. Exercise (5) on page 210 of [3]). (iii) For each $i = 1, 2, \cdots, k$ and any set C, $\lambda_y(C)$ is the conditional probability of the event $\{U_i\ \varepsilon\ C\}$ given $s(U_i) = y$ and any ω in Ω. Finally, (iv) "the conditional joint distribution of $U_1, U_2, \cdots$, and U_k given $s(U_i) = y_i$, $i = 1, 2, \cdots, k$, and ω is the product of the individual conditional distributions under the corresponding individual conditions", and the lemma follows. We omit the detailed verifications.

The reader should satisfy himself that (with a suitable definition of the sufficient statistic t in each case) the lemma applies to all standard cases of k independent samples of equal size. It is therefore likely to prove useful also in contexts other than the present one.

Now let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be a decision rule. Write

$$p_i^*(w) = \int_{R^{nk}} p_i(v)\ d\nu_w.$$

Since ν_w is a probability measure, it is clear from (2) that $0 \leqq p_i^*(w) \leqq 1$, $\sum_{i=1}^k p_i^*(w) \equiv 1$, so that $d^*(t(v)) = (p_1^*(t(v)), p_2^*(t(v)), \cdots, p_k^*(t(v))$ is a decision rule. It follows from Exercise (6) on page 211 of [3] that $p_i^*(w)$ is the conditional expectation of $p_i(V)$ given $t(V) = w$ and any ω in Ω; that is,

$$E[\chi_B(t(v))p_i(v) \mid \omega] = E[\chi_B(t(v))\, p_i^*(t(v)) \mid \omega]$$

for all subsets B of R^{mk} and all ω in Ω ($i = 1, 2, \cdots, k$). Taking $B = R^{mk}$ we see that $d^*(t(v))$ is equivalent to $d(v)$. Thus we have

THEOREM 1. *Corresponding to any decision rule $d(v)$ there exists an equivalent decision rule $d^*(t(v))$.*

Suppose now that $d(v)$ is an impartial decision rule. It is easy to see that in that case $d^*(t(v))$ must also be impartial. Consider the case $k = 2$. Then for any point (y_1, y_2) of R^{m2} we have

$$\begin{aligned} p_2^*(y_1, y_2) &= \int_{R^{n2}} p_2(v)\, d\nu_{(y_1, y_2)}(v) \\ &= \int_{R^n} \left\{ \int_{R^n} p_2(u_1, u_2)\, d\lambda_{y_1}(u_1) \right\} d\lambda_{y_2}(u_2) \\ &= \int_{R^n} \left\{ \int_{R^n} p_1(u_2, u_1)\, d\lambda_{y_1}(u_1) \right\} d\lambda_{y_2}(u_2) \\ &= \int_{R^n} \left\{ \int_{R^n} p_1(u_1, u_2)\, d\lambda_{y_1}(u_2) \right\} d\lambda_{y_2}(u_1) \\ &= \int_{R^{n2}} p_1(v)\, d\nu_{(y_2, y_1)}(v) = p_1^*(y_2, y_1) \end{aligned}$$

and the impartiality of $d^*(t(v))$ is proved. A parallel argument applies to the general case. Hence

THEOREM 2. *Corresponding to any impartial decision rule $d(v)$ there exists an equivalent impartial decision rule $d^*(t(v))$.*

We remind the reader that the $d^*(t(v))$ which we have constructed in terms of the given $d(v)$ is equivalent to $d(v)$ in virtue of the fact that $d^*(t(v))$ is the conditional expectation of $d(V)$ given $t(V) = t(v)$ and any ω in Ω. In many statistical applications, the loss incurred in adopting a particular decision $d^0 = (p_1^0, p_2^0, \cdots, p_k^0)$ when ω is the parameter point is of the form

$$l(\omega, d^0) = \sum_{i=1}^{k} f_i(\omega)p_i^0, \quad f_i(\omega) \geqq 0.$$

For each ω let $c(\omega, x)$ be a bounded convex function of x defined for $\min_i \{f_i(\omega)\} \leqq x \leqq \max_i \{f_i(\omega)\}$ and write $\psi(\omega, d^0) = c(\omega, l(\omega, d^0))$. Then ψ is a convex function of d^0 for each fixed ω. It follows from Lemma 3.1 of [4] that, irrespective of the particular weights $f_i(\omega)$ and particular function $c(\omega, x)$, we have

$$E[\psi(\omega, d^*(t(v))) \mid \omega] \leqq E[\psi(\omega, d(v)) \mid \omega] \text{ for all } \omega.$$

Our immediate purpose in stating this consequence of the relation between $d(v)$ and $d^*(t(v))$ will be served by noting the following easy corollaries of the general result: (i) The expected loss when using $d^*(t(v))$ always equals the expected loss when using $d(v)$, and (ii) the *variance* of the loss when using $d^*(t(v))$ never exceeds the variance when using $d(v)$. Now, the equivalence of $d^*(t(v))$ and $d(v)$ also implies (i), but results such as (ii) do not follow from equivalence alone. There is, therefore, a somewhat stronger justification than the one given by Theorems 1 and 2 for using decision rules which depend on the outcome v only through t.

We shall now give a useful representation of impartial decision rules. Let $\phi(u)$ be a real valued function on R^n and for any $v = (u_1, u_2, \cdots, u_k)$ set $\phi_i = \phi(u_i)$, $i = 1, 2, \cdots, k$. Let $\mathfrak{D}(\phi)$ be the class of all impartial decision rules which are based on $\phi_1, \phi_2, \cdots, \phi_k$ alone, that is, all impartial decision rules of the form $d(v) = (p_1(\phi_1, \phi_2, \cdots, \phi_k), p_2(\phi_1, \cdots, \phi_k), \cdots, p_k(\phi_1, \cdots, \phi_k))$. Since $\phi(u)$ is a given function, $\mathfrak{D}(\phi)$ will, in general, be a subclass of the class of all impartial decision rules, but may coincide with it. In any case, for given v, let $\phi_{(1)}, \phi_{(2)}, \cdots, \phi_{(k)}$ be the k (not necessarily distinct) numbers ϕ_i arranged in ascending order of magnitude and write

$$a_{ij} = \begin{cases} 1 & \text{if } \phi_i = \phi_{(j)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i, j = 1, 2, \cdots, k).$$

THEOREM 3. *A decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ is a member of $\mathfrak{D}(\phi)$ if and only if there exist functions $\lambda_j(z_1, z_2, \cdots, z_k)$, $j = 1, 2, \cdots, k$, such that*

$$0 \leqq \lambda_j \leqq 1, \qquad \sum_{j=1}^{k} \lambda_j \equiv 1$$

and such that for each $i = 1, 2, \cdots, k$

$$p_i(v) = \sum_{j=1}^{k} \frac{a_{ij}}{\sum_{i=1}^{k} a_{ij}} \lambda_j(\phi_{(1)}, \phi_{(2)}, \cdots, \phi_{(k)})$$

for all v.

The proof is by direct verification and is omitted.

3. An application. Let $\pi_1, \pi_2, \cdots, \pi_k$ be normal populations, π_i having an unknown mean m_i and variance σ^2. Write $\theta_i = (m_i, \sigma)$, $\omega = (\theta_1, \theta_2, \cdots, \theta_k)$, and let Ω be the set of all points ω which are regarded as being possible in the given case. Let $g_i(\omega)$, $i = 1, 2, \cdots, k$, be functions defined on Ω such that $m_i \leqq m_j$ implies $g_i \leqq g_j$, $i, j = 1, 2, \cdots, k$. Suppose that samples $u_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ of n independent observations are drawn from each of the populations π_i, giving the combined sample point $v = (u_1, u_2, \cdots, u_k)$. For any decision rule $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ and any ω in Ω let the expected loss, or risk, by given by

$$r(d \mid \omega) = \max_i \{g_i(\omega)\} - \sum_{i=1}^{k} g_i(\omega) \cdot E[p_i(v) \mid \omega]. \tag{6}$$

Regarded as a function of ω, $r(d \mid \omega)$ is called the *risk function* of $d(v)$. The problem is to construct (if possible) an impartial decision rule $d^*(v)$ such that $r(d^* \mid \omega)$ is as small as possible no matter what ω may be. We shall show that the problem has a solution which is independent of the functions g_i. We shall also describe two determinations of the functions g_i which seem to be of special interest.

For any v, set $\bar{x}_i = n^{-1} \sum_{j=1}^{n} x_{ij}$, $i = 1, 2, \cdots, k$, and $\bar{x}_{(1)} = \min\{\bar{x}_i\}$, $\bar{x}_{(k)} = \max\{\bar{x}_i\}$, and let $a(v)$ = number of $\bar{x}_i$'s which equal $\bar{x}_{(1)}$, $b(v)$ = number of $\bar{x}_i$'s which equal $\bar{x}_{(k)}$. (Of course, we have $\Pr(a(v) = b(v) = 1 \mid \omega) = 1$ for all ω.) Write

$$p_i^{(1)}(v) = \begin{cases} \dfrac{1}{a(v)} & \text{if } \bar{x}_i = \bar{x}_{(1)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, k),$$

$$p_i^{(k)}(v) = \begin{cases} \dfrac{1}{b(v)} & \text{if } \bar{x}_i = \bar{x}_{(k)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, k).$$

It is then clear that $d_1(v) = (p_1^{(1)}(v), p_2^{(1)}(v), \cdots, p_k^{(1)}(v))$ and $d_k(v) = (p_1^{(k)}(v), p_2^{(k)}(v), \cdots, p_k^{(k)}(v))$ are fixed impartial decision rules which depend on v only through $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$.

THEOREM 4. *Let $\mathfrak{D}$ be the class of all impartial decision rules $d(v)$. Then*

$$r(d_1 \mid \omega) = \sup_{d \varepsilon \mathfrak{D}} r(d \mid \omega), \qquad r(d_k \mid \omega) = \inf_{d \varepsilon \mathfrak{D}} r(d \mid \omega)$$

for all ω in Ω.

PROOF. Choose and fix an arbitrary impartial decision rule $d(v)$. Let $c > 0$ be any constant such that the subset $\Omega_c = \{\omega : \omega \varepsilon \Omega, \sigma = c\}$ is non-empty. Now, corresponding to each ω in Ω_c the probability density (with respect to n-dimensional Lebesgue measure) of the sample from π_i is of the form $h_c(u) g_i(\bar{x} : \omega)$, $i = 1, 2, \cdots, k$. Write $t(v) = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$. It follows from Theorem 2 that there exists an *impartial* decision rule based on $t(v)$ alone, say $d_c^*(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k)$, which is equivalent to $d(v)$ provided ω is restricted to Ω_c. From equivalence and (6), we have

$$r(d \mid \omega) = r(d_c^* \mid \omega) \tag{7}$$

for ω in Ω_c. Now, since for $i \neq j$ we have $\Pr(\bar{x}_i = \bar{x}_j \mid \omega) = 0$ for all ω, it follows that (with probability equal to one for all ω) the representation of impartial decision rules based on $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$ which is given by Theorem 3 coincides with the representation assumed in Theorem 1 of [1]. An application of this last theorem (cf. Example 1 in Section 6 of [1]) shows that

$$r(d_1 \mid \omega) \geqq r(d_c^* \mid \omega), \qquad r(d_k \mid \omega) \leqq r(d_c^* \mid \omega) \tag{8}$$

for all ω. It follows from (7) and (8) that $r(d_1 \mid \omega) \geqq r(d \mid \omega)$ and $r(d_k \mid \omega) \leqq r(d \mid \omega)$ for ω in Ω_c. Since both $d(v)$ and c are arbitrary, Theorem 4 is proved.

In conclusion, we describe two applications of Theorem 4. Suppose that v is the outcome of preliminary experiments on $\pi_1, \pi_2, \cdots, \pi_k$ and now it is desired to draw a total of N observations from the k populations in such a way that the mathematical expectation of the sum of the values obtained is as large as possible. Let $d(v) = (p_1(v), p_2(v), \cdots, p_k(v))$ be a suitable decision rule and suppose that $Np_i(v)$ observations are drawn from π_i, $i = 1, 2, \cdots, k$. Then the mathematical expectation of the sum of the values obtained is $N \sum_1^k m_i E[p_i(v) \mid \omega]$. Since the maximum of this quantity is $N \max \{m_1, m_2, \cdots, m_k\}$, the expected loss in using $d(v)$ may be taken to be

$$N\left[\max \{m_i\} - \sum_1^k m_i E[p_i(v) \mid \omega]\right]. \tag{9}$$

The expected loss is of the form (6), with $g_i(\omega) = Nm_i$ for $i = 1, 2, \cdots, k$. It follows that in the class of all impartial decision rules the uniformly best rule is to drawn an equal number of observations from populations π_i such that $\bar{x}_i = \max \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$ and none from the others.

Suppose now that it is desired to select one of the populations π_i, the object being to select a population, π_r say, such that $m_r = \max \{m_1, m_2, \cdots, m_k\}$. As pointed out in the introductory section, the statistician may then employ a suitable decision rule, say $d(v) = (p_1(v), \cdots, p_k(v))$, in the following way: given v, he performs a random experiment whose outcome ρ takes on the values $1, 2, \cdots, k$ with $\Pr(\rho = i) = p_i(v)$ $(i = 1, 2, \cdots, k)$, and selects π_ρ. Write $m_{(k)} = \max \{m_1, m_2, \cdots, m_k\}$, and set

$$g_i(\omega) = \begin{cases} 1 & \text{if } m_i = m_{(k)}, \\ 0 & \text{otherwise} \end{cases} \qquad (i = 1, 2, \cdots, k).$$

Then it is readily seen from (5) and (6) that with the present convention for the manner in which a decision rule $d(v)$ is to be used, we have

$$r(d \mid \omega) = \Pr \text{ (incorrect selection using } d(v) \mid \omega). \tag{10}$$

It follows from Theorem 4 that in the class of all impartial decision rules, the rule $d_k(v)$ which is to assign equal probabilities of selection to populations π_i such that $\bar{x}_i = \max \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$ and zero probabilities to the rest, minimizes the probability of an incorrect selection uniformly for all ω in Ω.

The reader is referred to [5] for an investigation from a more general viewpoint of the problem of minimizing (9) or (10) in the case $k = 2$. The discussion in [5] does not presuppose samples of equal size, and the class of all decision rules is taken into consideration.

REFERENCES

[1] R. R. Bahadur, "On a problem in the theory of k populations," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 362–375.

[2] H. A. SIMON, "Symmetric tests of the hypothesis that the mean of one normal population exceeds that of another," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 149–154.

[3] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Company, New York, 1950.

[4] J. R. HODGES AND E. L. LEHMANN, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[5] R. R. BAHADUR AND H. ROBBINS, "The problem of the greater mean," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 469–487.

[6] E. L. LEHMANN, "A general concept of unbiasedness," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 587–592.

SOME DISTRIBUTION-FREE TESTS FOR THE DIFFERENCE BETWEEN TWO EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTIONS

By E. F. Drion

Statistics Department, T.N.O., The Hague

1. Summary and introduction. It sometimes happens that of two empirical cumulative distribution curves (step curves) one lies entirely above the other, in other words that, except at both ends, they have no point in common. The problem then arises, what is the probability that this will happen when both are random samples from the same population. In this paper a partial answer will be given, based on the ingenious solution of André (as cited in the well known textbook of Bertrand [1] in the problem of the ballot and also in Chap. VIII, Sect. 5 of [7]). Moreover an analogous method will allow us to give an exact answer to the problem of the maximum difference between two empirical cumulative distribution functions of random samples from the same population, but only if both samples have the same size. Smirnov has given an asymptotic solution for the latter problem (cited by Feller [3], see also [2]).

Our result leads, by using the Stirling approximation for the factorials, to the asymptotic formula of Smirnov.

A comparison of numerical results of the exact formula and the asymptotic formula of Smirnov shows that at least in the case of equal samples, the probabilities calculated by the Smirnov formula have, for samples as small as 20, an error of less than 4% for probabilities 0.033 or more. (See also Massey [5], who has calculated the exact probabilities for equal samples by means of difference equations.)

2. Statement of the problem. Let a population P be given with an unknown continuous distribution function $F(x)$. From this population two random samples $x'_1 \cdots x'_{n_1}$ and $y'_1 \cdots y'_{n_2}$ are drawn. After ordering each sample from the smallest value to the greatest we shall call them $x_1 \cdots x_{n_1}$ and $y_1 \cdots y_{n_2}$. For each sample the empirical distribution-function (step-function) $F_1(x)$ or $F_2(y)$ is constructed:

$$
\begin{aligned}
F_1(x) &= 0, & x &< x_1, & F_2(y) &= 0, & y &< y_1, \\
F_1(x) &= \frac{i}{n_1}, & x_i &\leqq x < x_{i+1}, & F_2(y) &= \frac{j}{n_2}, & y_j &\leqq y < y_{j+1}, \\
F_1(x) &= 1, & x_{n_1} &\leqq x, & F_2(y) &= 1, & y_{n_2} &\leqq y.
\end{aligned}
$$

As we have assumed that the population has a continuous distribution-function, $\Pr(x_i = y_j) = 0$ for all sets of values of i and j; that is, the discontinuities of the two step-functions have, except for a probability zero, unequal abscissae.

Under these assumptions we ask for:

A. The probability that either $F_1(x) - F_2(x) < 0$ or $F_1(x) - F_2(x) > 0$ for all values of x between $\min(x_1, y_1)$ and $\max(x_{n_1}, y_{n_2})$ (boundaries not included).

B. The probability that $\max | F_1(x) - F_2(x) | \geqq d$.

We shall give a general solution of problem A both for the case that $n_1 = n_2$ and that the greatest common divisor of n_1 and n_2 equals one. For problem B a solution has only been found for the case that $n_1 = n_2$.

3. Graphical representation of two ordered samples. If we order the observations of both samples in one series according to their magnitude, so that we shall have a series of $n_1 + n_2$ terms of the form $x_1, x_2, y_1, x_3, y_2, \cdots, y_{n_2}$ say, then our problem A is equivalent to the following: What is the probability that, in a random series of n_1 x's and n_2 y's, the proportion of x's to y's from the first to the n-th term of the series (where n may have all values from 2 to $n_1 + n_2 - 1$ included) is, for each n, always smaller than n_1/n_2 or always larger than n_1/n_2.

That both problems are equivalent may be shown in this way. If the two series of observations are random samples from the same population, they may be considered as one sample of size $n_1 + n_2$, in which n_1 observations are marked x and n_2 are marked y. The marking of the observations does not depend (in random samples) on the result of the observations, so all orders of the x's and y's are equally probable.

To solve this problem we shall make use of a graphical representation of these series. Let the x's represent horizontal paces and the y's vertical paces, then all possible series will be represented by all possible routes joining the diagonal corners of a rectangular lattice of sides n_1 and n_2. Those routes which have no common point (except the end-points), with the diagonal of our rectangle, represent series where the proportion of x's to y's is either always larger than n_1/n_2 or always smaller.

As an illustration we shall give the step-curves and the routes in the lattice for two series, in one of which the step-curves do not have a point in common, (and where, therefore, the route in the lattice lies entirely at one side of the diagonal) while in the other the step-curves intersect[1]. The sequence of ordered samples in Fig. 1 is (roman type denoting x's and italic denoting y's) 2.0, 2.3, 2.4, *2.6*, 2.7, 2.9, 3.0, *3.1*, *3.3*, 3.4, *3.6*, 3.8, *4.1*. The sequence in Fig. 2 is 2.0, 2.3, *2.5*, 2.6, *2.8*, 2.9, 3.1, *3.2*, 3.4, 3.5, *3.6*, *4.3*, *4.5*.

The number of all possible routes from O to P is $\binom{n_1 + n_2}{n_1} = T$. We shall now calculate the number A of all routes A[2] from O to P lying below the diagonal OP. The fraction A/T gives then the probability that of two empirical cumulative distribution curves of samples from one population the second lies entirely above the first. As each of the samples may be chosen as the first, the probability of no intersections of the step curves will be $2A/T$.

---

[1] It will be clear that if the paces in both directions have unit length, the route divides the rectangle in two parts of which the area's are respectively U and $n_1 n_2 - U$, where U is the statistic defined by Mann and Whitney for the test of Wilcoxon [4].

[2] We use A as well to indicate a route lying entirely to the right of the diagonal as to indicate the number of these routes.

The number A of routes lying below the diagonal OP depends on the number of lattice-points on OP that is to say, on the greatest common divisor of n_1 and n_2. If $n_1 = n_2 = n$ all routes reaching the diagonal will reach it in a lattice-point, as no route can intersect the diagonal except in a lattice-point. If n_1 and n_2 are coprime there are no lattice-points on the diagonal (except the endpoints O and P), while if n_1 and n_2 ($n_1 \neq n_2$) have a greatest common divisor $d > 1$, there

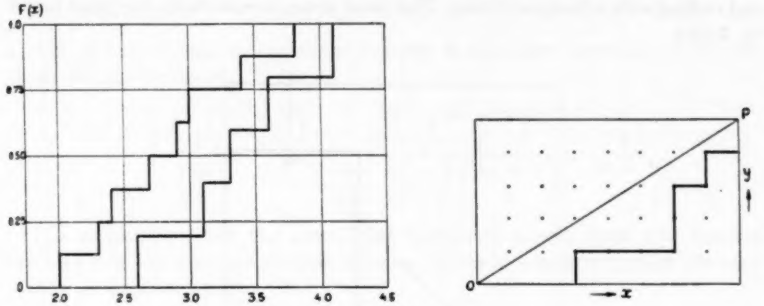

FIG. 1

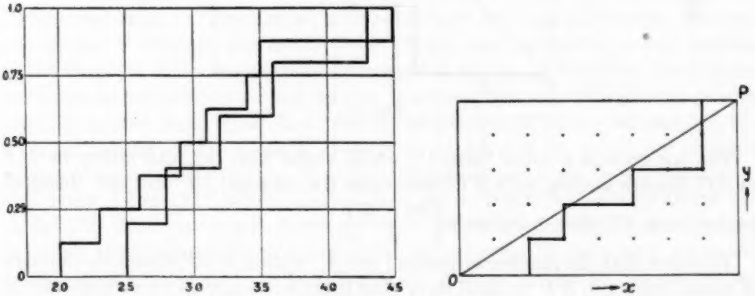

FIG. 2

are $d - 1$ lattice-points on the diagonal between O and P; so on $n_1 - d$ points a vertical route section and on $n_2 - d$ points a horizontal route section can intersect the diagonal outside a lattice point.

So the lattice-points available for a route under the diagonal OP is relatively to the total number lattice-points highest if n_1 and n_2 are coprime and lowest if $n_1 = n_2$. It stands to reason that the number of routes A is in the first case higher than in the second case. This we shall prove. For the intermediary case (greatest common divisor d of n_1 and $n_2 > 1$) we shall prove that the number of routes A relative to the total number of routes T is always less than when n_1 and n_2 are coprime. Probably this number is always higher than when $n_1 = n_2$. But we were not able to prove it.

4. Determination of the number of routes A in the case $n_1 = n_2 = n$. In this case (Fig. 3) the lattice is a square with $(n + 1)^2$ points. We shall not determine the number of routes A directly, but first we shall determine the number of routes that start with a horizontal step OR (and so could belong to the class A) having at least one point in common with the diagonal. It will be proved that this number equals twice the number of routes starting with a horizontal step and ending with a horizontal step. The proof given is essentially the proof found by André.

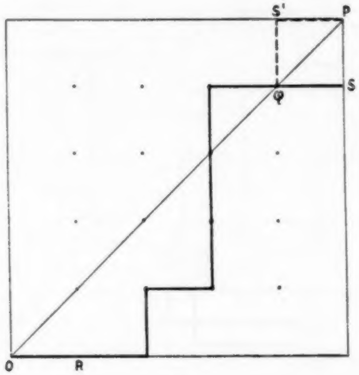

Fig. 3

The last step of a route "not-A", which starts with OR, can either be $S'P$ or SP. Routes ending with $S'P$ must cross the diagonal OP and are therefore routes "not-A"; their number is $\binom{2n-2}{n}$.

To prove that the number of routes "not-A" ending in SP equals the number of routes ending in $S'P$ we shall show that there exists a one-one correspondence between the routes "not-A" ending in SP and the routes ending in $S'P$. A route "not-A" like ORQSP can be transformed in a route ending in $S'P$ by rotating the part QSP about OP to $QS'P$. Here the point Q is the last point on the route before P that lies on the diagonal OP; each route "not-A" ending in SP can therefore be transformed in one way only in a route ending in $S'P$. On the other hand each route beginning with OR and ending with $S'P$ will cross at least once the diagonal OP. By rotating about the diagonal OP that part of the route, which lies between P and the point Q where it reaches for the first time OP, it will be transformed in a route "not-A" ending in SP. This route "not-A" ending in SP is also uniquely determined by the route ending in $S'P$. So we have proved the one-one correspondence between the routes "not-A" ending in SP and the routes "not-A" ending in $S'P$. The total number of routes "not-A" starting with OR is therefore $2\binom{2n-2}{n}$.

The total number of routes starting with OR is $\binom{2n-1}{n}$ therefore the num- of routes A is

$$\binom{2n-1}{n} - 2\binom{2n-2}{n} = \left(\frac{2n-1}{n-1} - 2\right)\binom{2n-2}{n} = \frac{1}{n-1}\binom{2n-2}{n}.$$

The total number T of routes from O to P is $\binom{2n}{n}$. So the probability that a route chosen at random lies either entirely to the right or entirely to the left of the diagonal equals

$$\frac{2 \times \frac{1}{n-1}\binom{2n-2}{n}}{\binom{2n}{n}} = \frac{\frac{2}{n-1}\binom{2n-2}{n}}{\frac{2n}{n}\cdot\frac{2n-1}{n-1}\binom{2n-2}{n}} = \frac{1}{2n-1}.$$

The probability that the cumulative frequency curves from two random samples n of the same population have no points in common (except the endpoints) is therefore $(1/2n - 1)$.

5. Determination of the number of routes A in the case n_1 and n_2 coprime. In this case (Fig. 4) there are no lattice-points on the diagonal except the endpoints, and if through any lattice-point (except the endpoints) a line parallel to the diagonal is drawn no other lattice-point will lie on this line; for if there were two lattice-points x_1y_1 and x_2y_2 on this line, then the triangle with angles (x_1y_1), (x_2y_2) and (x_2y_1) would be similar to the triangle $(0, 0)$; (n_1, n_2) and $(n_1, 0)$; so $(y_2 - y_1)/(x_2 - x_1) = n_2/n_1$, where $(y_2 - y_1)$ and $(x_2 - x_1)$ are integers smaller than n_2 respectively n_1. But this is impossible, as n_1 and n_2 are coprime.

A route A like OQP passes through $n_1 + n_2 - 1$ lattice-points (O and P excluded). If this route is cut in any of those lattice-points (like Q) and the two parts are interchanged the new route will not be a route A, that is to say it will not lie entirely to the right of the diagonal OP. For the angle PQC' is greater than the angle POC, so that if Q is placed in O then P will lie in a point Q' to the left of OP. Furthermore a straight line through Q' parallel to OP will not intersect anywhere the polygon $OQ'P$; the part OQ' is not intersected because OP does not intersect the part QP of the original line and $Q'P$ is not intersected because OP does not intersect OQ (for OQP is a route A that is, by definition a route, not intersected by OP). If we cut the route $OQ'P$ in Q', (which point is uniquely determined as being the first point lying on a line parallel to OP moved from D to P) and interchange the two parts OQ' and $Q'P$, the original route OQP will be reconstructed. On each route $OQ'P$ which passes through at least one lattice-point Q' at the left-hand side of OP and only on these routes, one, and only one, point Q' can be found, therefore a route $OQ'P$ (not-A) gives after transformation only one route $OQP(A)$. On the other hand, two different cuts of a route A will give after transformation two different routes, because if the coordinates of the section-points be (x_1, y_1) respectively (x_2, y_2), the coordinates of the images Q' respectively Q_2' of O and P will be $(n_1 - x_1, n_2 - y_1)$ respectively $(n_1 - x_2, n_2 - y_2)$, which points are different. As each route "not-A" has only one point Q', two routes with different points Q' are different. It is also impossible that two *different* routes "A" give after section the same route "not-A", because the transformation of a route "not-A" to a route "A" is unique. As all routes lie either entirely to the right of OP (are routes "A") or have at least one point to the right of OP, and as each route A gives by the $(n_1 + n_2 - 1)$ possible cuts $(n_1 + n_2 - 1)$ different routes "not-A", the total number T of routes from O to P equals $A + (n_1 + n_2 - 1)A = (n_1 + n_2)A$. Therefore the probability that a randomly chosen route is a route A equals $1/(n_1 + n_2)$. The probability that two empirical cumulative distribution-curves, from two samples of size n_1 and n_2 (n_1 and n_2 coprime) from the same population do not intersect, is therefore $2/(n_1 + n_2)$.

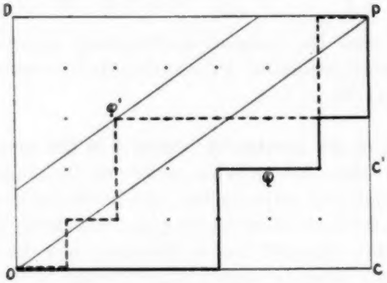

FIG. 4

6. Determination of the number of routes A in the case n_1 and n_2 have a common divisor greater than 1. If n_1 and n_2 $(n_1 \neq n_2)$ have a greatest common divisor $d > 1$, there are $d - 1$ lattice-points on the diagonal (except the end-points). In this case the routes "not-A" can be divided into two groups: "not-A_1", routes which pass through at least one lattice-point at the lefthandside of the diagonal, and "not-A_2", routes which pass through one or more lattice-points on the diagonal but do not pass through a lattice-point at the lefthandside of the diagonal. A cut followed by an interchange of the two halves of a route "A" will transform it into a route "not-A_1". A cut followed by an interchange of the two halves of a route "not-A_2" will transform it either into a route "not-A_1", or into another or the same route "not-A_2" (if the cut falls on the diagonal). So the total number of routes is $(n_1 + n_2 - 1)\, A + A +$ routes "not-A_2" $+$ routes "not-A_1", resulting from cuts in routes "not-A_2", $= (n_1 + n_2)\, A + x$. The number of routes A is therefore less than the $1/(n_1 + n_2)$th part of all the routes.

It may seem rather strange that the probability in the case n_1, n_2 coprime is about twice the probability in the case $n_1 = n_2$. This discontinuity is of course caused by the fact that in the case $n_1 = n_2$ both distribution-curves may have one or more points in common (except the endpoints) without crossing each other (in other words the graph may meet the diagonal without crossing it). In the case that n_1 and n_2 are coprime this is impossible.

If in the case $n_1 = n_2$ we seek the probability that either $F_1(x) - F_2(x) \leqq 0$ or $F_1(x) - F_2(x) \geqq 0$ (instead of $F_1(x) - F_2(x) < 0$ or $F_1(x) - F_2(x) > 0$) it will be found (by applying formula (1) of the next section with $h = 1$) that this probability is $2/(n + 1)$. Therefore under these conditions the probability for $n_1 = n_2$ is about twice as large as for n_1 coprime to n_2.

In consequence no direct statistical test can be based on these results, as one of the referees remarked. However should one in an investigation find that of two empirical distribution-curves, one lies entirely above the other, the formulas given above enable one to calculate the probability that such a result is caused by random sampling fluctuations.

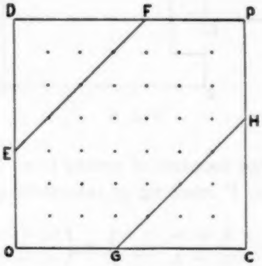

FIG. 5

7. Probability that the maximum difference of two empirical distribution curves from two samples of size n from one population is at least h/n. To solve this problem (exact solution of the problem of Smirnov in the case of equal samples) we shall again use the representation of our two samples by the lattice OCPD (Fig. 5). As $n_1 = n_2 = n$, this lattice is a square. All routes from O to P that reach a point on one of the lines EF, GH or on both lines, or that intersect one or both of these lines represent pairs of samples, where for some value x the maximum difference $| F_1(x) - F_2(x) |$ is at least $OE/PC = OE/n$.

To solve this problem we need the following lemma.

LEMMA. *The number of routes in a rectangular lattice with sides n_1 and n_2, such that somewhere the number of vertical paces y exceeds the number of horizontal paces x by at least h is* $\binom{n_1 + n_2}{n_1 + h}$. (An algebraic solution of this problem is given in Whitworth Proposition XXIX [8]. We shall give here a geometrical solution that can be extended to the problem of Smirnov.)

PROOF. All routes, such that somewhere the number of vertical paces exceeds the number of horizontal paces by at least h, are routes that reach or intersect a line EF, which makes an angle of 45° with DO (fig. 6).

We shall cut a route, such as OGP' that somewhere reaches the line EF, in the point G where it reaches this line for the first time. The part OG is reflected about EF to $O'G$; the part GP' is left in its place.

A route from O to P', reaching or intersecting EF, may thus be transformed in one way in a route from O' to P'. As we may transform the route $O'P'$ back to the route OGP' by cutting it in the first place where it reaches EF, and reflecting $O'G$ about EF, we see that there is a one to one correspondence between the routes from O to P' reaching or intersecting EF and the routes ($O'GP'$) from O' to P'. If the sides of the lattice measure n_1 ($= OC'$) and n_2 ($= C'P'$) and if OE

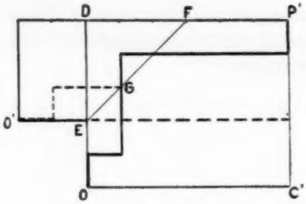

FIG. 6

$= O'E$ measures h, then the number of routes from O' to P and therefore the number of routes from O to P reaching or intersecting the line EF is

$$\binom{n_1 + h + n_2 - h}{n_1 + h} = \binom{n_1 + n_2}{n_1 + h} \tag{1}$$

7.1. *Classification of routes OP representing empirical distributions where* max $| F_1(x) - F_2(x) | \geqq h/n$. The solution of the problem of Smirnov is, even in the case of equal samples, rather complicated, while the empirical distribution curves $F_1(x)$ and $F_2(x)$ may intersect more than once. Therefore it is possible that there are one or more values of x such that $F_1(x) - F_2(x) \geqq h/n$, and in the same pair of samples other values of x, such that $F_2(x) - F_1(x) \leqq h/n$. In other words, the route may intersect or touch both lines $a = EF$ and $b = GH$. We may classify the routes which touch or cross either one or both of the lines a and b in the following way (Fig. 5):

A. routes touching or crossing only a,

B. routes touching or crossing only b,

C. routes touching or crossing first one or more times a and afterwards touching or crossing b (after having touched or crossed b, these routes may also touch or cross a again).

D. routes touching or crossing first one or more times b and afterwards touching or crossing a (after having touched or crossed a, these routes may also touch or cross b again). The letters A, B, C and D will also be used for the number of routes of these categories. In the same way we will use the letters a and b also for the number of all routes that cross a respectively b, whether they cross b

(respectively a) or not. By reasons of symmetry it is clear that $a = b$, $A = B$ and $C = D$. Furthermore $a = A + C + D$, $b = B + C + D$. The total number of routes touching or crossing either a or b or both is $A + B + C + D = a + b - (C + D) = 2(a - D)$.

The number of routes a is given by our lemma viz. $\binom{2n}{n-h}$, so we have only to find the number of routes D.

7.2. *Calculation of the number of routes D.* The number of routes D may be ound by a repeated application of the device used for calculating a (Fig. 7).

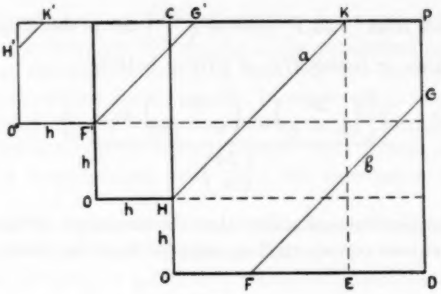

FIG. 7

We rotate the rectangle $OEKC$ about HK, leaving that part of the route from O to P unchanged which begins at the point where this route touches or crosses for the first time HK. After the transformation the routes D are the routes from O' to P which touch or cross $F'G'$, *without having first crossed the line HK*. The lengths of the sides of the rectangle $O'P$ (indicating this rectangle by the ends of one of the diagonals) are $n - h$ and $n + h$. To determine the total number of routes touching or crossing $F'G'$, we rotate the rectangle $O'G'$ about $F'G'$, leaving unchanged that part of the route from O' to P which begins at the point where this route touches or crosses for the first time $F'G'$. The new rectangle will have sides $n - 2h$ and $n + 2h$. The total number of routes in the rectangle $O''P$ is $\binom{2n}{n-2h}$, this is therefore the total number of the routes in the rectangle $O'P$ which touch or cross the line $F'G'$. To get the number of routes D we must subtract from this number $\binom{2n}{n-2h}$ the number of routes in $O''P$ touching or crossing the image $H'K'$ of HK in $O''P$, *without* having touched or crossed $F'G'$. By rotating the rectangle $O''K'$ about $H'K'$, we get a new rectangle with sides $n - 3h$ and $n + 3h$. The total number of routes in this rectangle is $\binom{2n}{n-3h}$; the sought number of routes to subtract from $\binom{2n}{n-2h}$ among these are the routes which do not touch or cross the image of $F'G'$, which can be determined by repeating the process of rotating. The law for the determination of the number of routes will be clear. Their number is: $\binom{2n}{n-2h} - \binom{2n}{n-3h} + \binom{2n}{n-4h} - \cdots$, the series being continued as long as $n - kh \geqq 0$.

The total number of routes which cross either one or both lines HK and FG is therefore

$$2\left[\binom{2n}{n-h} - \binom{2n}{n-2h} + \binom{2n}{n-3h} \cdots\right].$$

As all the routes from O to P number $\binom{2n}{n}$ the probability that a random chosen route touches or crosses HK or FIG or both is

$$\frac{2\left[\binom{2n}{n-h} - \binom{2n}{n-2h} + \cdots\right]}{\binom{2n}{n}}.$$

This is therefore also the probability that the maximum difference of the cumulative frequency-curves of two random samples from the same population is at least h/n.

TABLE I

n	d	$h = nd$	P	
			Exact	Smirnov
20	.25	5	.5713	.5596
20	.40	8	.0811	.0815
20	.45	9	.0335	.0349
20	.50	10	.0123	.0135
50	.16	8	.5487	.5441
50	.24	12	.1124	.1123
50	.28	14	.0392	.0396
50	.32	16	.0115	.0120
100	.12	12	.4695	.4676
100	.17	17	.1112	.1112
100	.19	19	.0539	.0541
100	.23	23	.0099	.0101

8. Some numerical results. We have calculated the probability P that $\max | F_1(x) - F_2(x) | \geqq d$ for samples of size $n = 20, 50, 100$ and for values of d such that $P \sim 0.50$, 0.10, 0.05 and 0.01 by means of the exact formula and by the asymptotic formula of Smirnov. The results are given in table I.

The figures, given in the last column, were found by linear interpolation in the table of Smirnov [6].

With equal-sized samples the asymptotic formula of Smirnov gives very satisfactory results even for samples of 20. We suspect, however, that when the samples are of unequal size the agreement will be less satisfactory especially if n_1 and n_2 are coprime, because in this case there is only one lattice point on HK and FG, which must in this case be parallel to the diagonal OP (c.f. Fig. 7).

9. Concluding remarks. (a) The probabilities given above are based on the assumption that the distribution-functions of the population are continuous. In practice almost all distribution-functions, however, are discontinuous, owing to the limited accuracy of our measurements. In other words, in practice we work always with grouped data, although the classes may be so small, that in no class falls more than one observation and often none. Nevertheless, when the number of observations is large enough, more than one observation will be found in several classes.

Let the width of the classes be h, so that the values of $F(x)$ (i.e. the cumulative experimental distribution-function) are only known for $x = hg$ (with g an integer between $g_a = [x_1/h]$ and $g_b = [x_n/h] + 1$, where $[x/h]$ denotes the integer part of x/h). If of two ungrouped samples, $x_1, \cdots, x_{n_1}$ and $y_1, \cdots, y_{n_2}$, the cumulative experimental distribution curve of the y's lies entirely to the righthand side of that of the x's, i.e. if $F_1(x) > F_2(x)$, $a_1 = \min(x_1, y_1) \leqq x \leqq \max(x_{n_1}, y_{n_2}) = a_2$ then, after grouping,

$$F_1(hg_i) > F_2(hg_i), \qquad g_1 = \min\left(\left[\frac{x_1}{h}\right] + 1, \left[\frac{y_1}{h}\right] + 1\right) \leqq g_i$$
$$\leqq \max\left(\left[\frac{x_{n_1}}{h}\right], \left[\frac{y_{n_2}}{h}\right]\right) = g_n.$$

But the converse needs not hold. Therefore the probability that $F_1(hg_i) > F_2(hg_i)$ for all values of g_i between g_1 and g_n (g_1 and g_n included) is greater than or equal to the probability that $F_1(x) > F_2(x)$ for all values of x between a_1 and a_2 (a_1 and a_2 included).

If however $F_1(hg_i) > F_2(hg_{i+1})$, $g_1 \leqq g_i < g_n - 1$, then $F_1(x) > F_2(x)$, $a_1 \leqq x \leqq a_2$, although the converse needs not hold. Therefore the probability that $F_1(hg_i) > F_2(hg_{i+1})$ for all values of g_i between g_1 and $g_n - 1$ is less than or equal to the probability that $F_1(x) > F_2(x)$ for all values of x between a_1 and a_2 (a_1 and a_2 included).

From this last result the following conclusion may be drawn: the probability that in two grouped random samples from the same population the cumulative experimental frequencies of one of the samples is higher at all class boundaries (which are the only values of the variate for which the cumulative frequency is known) than the cumulative frequencies of the other sample at the class boundaries of the next higher class is less than or equal to the formulae given in Sections 4 and 5.

(b) The formula given in Section 7 for the Smirnov test applies to the two-sided test. In the case we are only interested in deviations in one direction the formula is much simpler. With equal-sized samples from the same population the probability that $F_1(x) - F_2(x) > d/n$ is $\binom{2n}{n-d} \Big/ \binom{2n}{n}$.

REFERENCES

[1] J. Bertrand, *Calcul des Probabilités*, Paris, 1907.

[2] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[3] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–189.

[4] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[5] Frank J. Massey, "The distribution of the maximum deviation between two sample cumulative step functions" *Annals of Math. Stat.*, Vol. 22 (1951), pp. 125–128.

[6] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[7] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw Hill Book Co., New York 1937.

[8] W. A. Whitworth, *Choice and Chance*, Cambridge, 1901.

CONFIDENCE BOUNDS FOR A SET OF MEANS

By D. A. S. Fraser

University of Toronto

1. Summary. Professor John Tukey suggested the following two problems to the author: given that $X_1, X_2, \cdots, X_n$ are normally and independently distributed with unknown means $\mu_1, \mu_2, \cdots, \mu_n$ and given variance σ^2;

Problem A: *Find a β-level confidence interval of the form*

$$g(x_1, \cdots, x_n) \geqq \mu_1, \cdots, \mu_n \geqq -\infty.$$

Problem B: *Find a β-level confidence interval of the form*

$$g(x_1, \cdots, x_n) \geqq \mu_1, \cdots, \mu_n \geqq h(x_1, \cdots, x_n).$$

The main result of this paper is the nonexistence of intervals satisfying mild regularity conditions and having an exact confidence level (unless $n = 1$ or $\beta = 0, 1$). However for each problem an interval is given for which the confidence level is greater than or equal to β (formulas (2.1), (4.1)); these intervals are apparently shorter than those previously used in practice. Also the procedure for obtaining any interval with at least β confidence is described.

Some results are discussed for distributions other than the normal.

2. Introduction to Problem A.

2.1. *Normal distributions.* If $X_1, \cdots, X_n$ are normally and independently distributed with known variance σ^2 and unknown means $\mu_1, \cdots, \mu_n$, then Problem A is to find an upper β-level confidence bound for the set $\{\mu_1, \cdots, \mu_n\}$; that is, to find a function $g(x_1, \cdots, x_n)$ such that $\Pr\{g(X_1, \cdots, X_n) \geqq \max \mu_i\} \geqq \beta$ for all $\mu_1, \cdots, \mu_n$.

One approach to this problem is to look for exact β-level confidence bounds: the above condition on the function $g(x_1, \cdots, x_n)$ is replaced by $\Pr\{g(X_1, \cdots, X_n) \geqq \max \mu_i\} = \beta$ for all $\mu_1, \cdots, \mu_n$. This more restrictive condition in a confidence region problem is of course analogous to the requirement of similarity in the theory of hypothesis testing.

In Section 3 Problem A is analyzed but attention is confined to measurable functions $g(x_1, \cdots, x_n)$ which satisfy two mild restrictions. These restrictions are given by the following assumptions concerning the function $g(x_1, \cdots, x_n)$.

Assumption 2.1. *For all $x_1, \cdots, x_n$, $g(x_1 + \delta, \cdots, x_n + \delta)$ is a monotone nondecreasing function of δ.*

Assumption 2.2. *If $x_j = \max x_i$ $(i = 1, \cdots, n)$, then $g(x_1, \cdots, x_n)$ satisfies*

$$g(x_1, \cdots, x_n) \leqq g(x_1, \cdots, x_{j-1}, x_j + \delta, x_{j+1}, \cdots, x_n)$$

for all $x_1, \cdots, x_n$ and for any positive δ.

The second assumption seems reasonable since a bound would certainly be suspect if it were smaller for 27.2, 25.5, 26.3, 27.8 than for 27.2, 25.5, 26.3, 27.5.

It is then proved by Theorem 1 that there does not exist an exact β-level confidence bound which satisfies these two assumptions.

As a by-product of Theorem 1 a bound having at least β confidence is obtained; it is

$$g(x_1, \cdots, x_n) = \max x_i + N_{1-\beta}\,\sigma, \tag{2.1}$$

where $N_{1-\beta}$ is the $1 - \beta$ point of the unit normal; that is,

$$\frac{1}{\sqrt{2\pi}} \int_{N_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\, dx = \alpha.$$

The optimum properties of this bound will be discussed in a later paper.

The above bound, however, is not the only confidence bound. In Section 3 a procedure is given for constructing bounds having at least β confidence. For this it is convenient to restrict attention to bounds satisfying a more restrictive version of Assumption 2.1. This assumption 2.1* is obtained by applying the principle of cogredience to the problem using the transformations $x_i' = x_i + C$, $i = 1, \cdots, n$, for all C.

ASSUMPTION 2.1*. *The function $g(x_1, \cdots, x_n)$ satisfies the equality*

$$g(x_1 + \delta, \cdots, x_n + \delta) = g(x_1, \cdots, x_n) + \delta$$

for all $x_1, \cdots, x_n, \delta$.

2.2. *The general problem.* The problem as described above is a particular case of the following: given $X_1, \cdots, X_n$ are independently distributed with probability density functions $f(x - \mu_1), \cdots, f(x - \mu_n)$, find a β-level confidence bound for the set $\{\mu_1, \cdots, \mu_n\}$. Theorem 2 shows that if $f(x - \mu)$ satisfies a condition of bounded completeness, then exact β-level bounds do not exist. A bound having at least β confidence can of course always be obtained by adding to $\max x_i$ the $1 - \beta$ point of the distribution $f(x)$ (that is, with $\mu = 0$).

3. Analysis of Problem A.

3.1. *Characteristic function of a confidence bound.* We define a characteristic function for the bound $g(x_1, \cdots, x_n)$ as follows:

$$\begin{aligned} \phi_\theta(x_1, \cdots, x_n) &= 1, \qquad & g(x_1, \cdots, x_n) &\geqq \theta, \\ &= 0, & g(x_1, \cdots, x_n) &< \theta. \end{aligned} \tag{3.1}$$

From assumption 2.1 we can infer that $\phi_\theta(x_1 + \delta, \cdots, x_n + \delta)$ is a monotone nondecreasing function of δ.

To derive conditions on $\varphi_\theta(x_1, \cdots, x_n)$ from Assumption 2.2 we first define disjoint sets $S_1, \cdots, S_n$ which cover R^n except for a set of measure zero:

$$S_i = \{(x_1, \cdots, x_n) \mid x_i > \max_{j \neq i} x_j\}. \tag{3.2}$$

The second assumption insures that for points $(x_1, \cdots, x_n) \varepsilon S_i$ $\phi_\theta(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ is a monotone nondecreasing function of x_i.

3.2. *Theorem for normal variables.* To prove Theorem 1 we shall need the following

LEMMA 1. *If $Y_1, \cdots, Y_n$ are normally and independently distributed with means $\mu_1, \cdots, \mu_n$ and unit variances, then the set of densities corresponding to all $(\mu_1, \cdots, \mu_n) \varepsilon [-\infty, 0]^n$ is boundedly complete; that is,*

$$E\{\phi(Y_1, \cdots, Y_n)\} \equiv 0, \quad (\mu_1, \cdots, \mu_n) \varepsilon [-\infty, 0]^n,$$

and

$$|\phi(y_1, \cdots, y_n)| < M$$

imply

$$\phi(y_1, \cdots, y_n) = 0$$

almost everywhere.

PROOF: The above set of distributions is complete; see Lehmann and Scheffé [1]. Since completeness implies bounded completeness, the lemma follows.

THEOREM 1. *If $X_1, \cdots, X_n$ are normal and independent with means $\mu_1, \cdots, \mu_n$ and variance σ^2, there does not exist (unless $\beta = 0, 1$, or $n = 1$) a measurable function $g(x_1, \cdots, x_n)$, satisfying assumptions* 2.1 and 2.2, *which is an exact β confidence bound for $\max \mu_i$.*

PROOF: Without loss of generality let $\sigma^2 = 1$. We consider a measurable function $g(x_1, \cdots, x_n)$ satisfying assumptions 2.1 and 2.2 and, assuming that g is an exact β-level confidence bound, we shall find that a contradiction results.

We have

$$\begin{aligned} \beta &= \Pr\{g(X_1, \cdots, X_n) \geqq \max \mu_i\} \\ &= E\{\phi_\theta(X_1, \cdots, X_n) \mid \max \mu_i = \theta\} \\ &= E\{\beta_\theta^{(i)}(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) \mid \max_{j \neq i} \mu_j < \theta\}, \end{aligned} \tag{3.3}$$

where

$$\beta_\theta^{(i)}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi_\theta(x_1, \cdots, x_n) e^{-\frac{1}{2}(x_i - \theta)^2}\, dx_i. \tag{3.4}$$

We now derive conditions on the function $\beta_\theta^{(i)}$ and for simplicity let $\theta = 0$. From the expression above it is seen that

$$E\{\beta_0^{(i)}(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n) - \beta \mid \max_{j \neq i} \mu_j < 0\} = 0;$$

hence from Lemma 1, we conclude that

$$\beta_0^{(i)}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) = \beta \tag{3.5}$$

almost everywhere.

Using the above condition on $\beta_0^{(i)}$, we obtain conditions on the function $\phi_0(x_1, \cdots, x_n)$.

$$
\begin{aligned}
\beta &= \beta_0^{(i)}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) \qquad \text{almost everywhere} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi_0(x_1, \cdots, x_n)\, e^{-\frac{1}{2}x_i^2}\, dx_i .
\end{aligned} \tag{3.6}
$$

Consider fixed $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$ (not of course belonging to the exceptional set of measure zero for which the equality (3.5) might not hold). For $x_i > \max_{j \neq i} x_j$, $\phi_0(x_1, \cdots, x_n)$ is a monotone function; and since it is a characteristic function it will have the following form

$$
\begin{aligned}
\phi_0\,(x_1, \cdots, x_n) &= 0, \qquad \max_{j \neq i} x_j < x_i < u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n), \\
&= 1, \qquad u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) < x_i < \infty.
\end{aligned} \tag{3.7}
$$

$u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ is taken to be the value of x_i at which $\phi_0(x_1, \cdots, x_n)$ jumps from 0 to 1 or $\max_{j \neq i} x_j$, whichever is larger. Using the function $u(x_1, \cdots, x_n)$, we obtain

$$
\begin{aligned}
\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\max_{j \neq i} x_j} \phi_0(x_1, \cdots, x_n) e^{-\frac{1}{2}x_i^2}\, dx_i \\
+ \frac{1}{\sqrt{2\pi}} \int_{u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)}^{\infty} e^{-\frac{1}{2}x_i^2}\, dx_i .
\end{aligned} \tag{3.8}
$$

However, since

$$
\begin{aligned}
0 &\leqq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\max_{j \neq i} x_j} \phi_0(x_1, \cdots, x_n) e^{-\frac{1}{2}x_i^2}\, dx_i , \\
&\leqq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\max_{j \neq i} x_j} e^{-\frac{1}{2}x_i^2}\, dx_i \\
&= \Pr(X_i \leqq \max_{j \neq i} x_j),
\end{aligned}
$$

then

$$
N_\beta \leqq u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) \leqq N_{\beta - P(\max_{j \neq i} x_j)}, \tag{3.9}
$$

where

$$
P(\max_{j \neq i} x_j) = \Pr\{X_i \leqq \max_{j \neq i} x_j\}.
$$

The inequality on $u_i(x_1, \cdots, x_n)$ implies that $\phi_0(x_1, \cdots, x_n)$ is equal to zero for almost all points in S_i having $x_i < N_\beta$. This is true for all i; hence $\phi_0(x_1, \cdots, x_n) = 0$ if $\max x_j < N_\beta$. Consider now $(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ having $\max_{j \neq i} x_j < N_\beta$; in expression (3.8), the first integral vanishes leaving

$$
\beta = \frac{1}{\sqrt{2\pi}} \int_{u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)}^{\infty} e^{-\frac{1}{2}x_i^2}\, dx_i .
$$

Therefore

$$u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) = N_\beta, \qquad \max_{j \neq i} x_j < N_\beta.$$

From the above equality on $u(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$, we obtain the following conditions on $\phi_0(x_1, \cdots, x_n)$:

$$\begin{aligned} \phi_0(x_1, \cdots, x_n) &= 0, & &\text{if } \max_1^n x_i < N_\beta, \\ &= 1, & &\text{if exactly one } x \text{ is larger than } N_\beta. \end{aligned}$$

But since $\phi_0(x_1 + \delta, \cdots, x_n + \delta)$ is monotone in δ, we have

$$\begin{aligned} \phi_0(x_1, \cdots, x_n) &= 0, & &\text{if } \max x_1 < N_\beta, \\ &= 1, & &\text{if } \max x_i > N_\beta. \end{aligned} \tag{3.10}$$

Therefore

$$\begin{aligned} g(x_1, \cdots, x_n) &< 0, & &\text{if } \max x_i < N_\beta, \\ &\geqq 0, & &\text{if } \max x_i > N_\beta. \end{aligned}$$

Similarly

$$\begin{aligned} g(x_1, \cdots, x_n) &< \theta, & &\text{if } \max x_i < N_\beta + \theta, \\ &\geqq \theta, & &\text{if } \max x_i > N_\beta + \theta. \end{aligned} \tag{3.11}$$

This completely determines $g(x_1, \cdots, x_n)$;

$$\begin{aligned} g(x_1, \cdots, x_n) &= \max x_i - N_\beta \\ &= \max x_i + N_{1-\beta}. \end{aligned} \tag{3.12}$$

However, contrary to our original assumption, this function $g(x_1, \cdots, x_n)$ is not an exact β-level confidence bound unless $\beta = 0, 1$, or $n = 1$. For consider $\beta_0^{(1)}(x_2, \cdots, x_n)$; (3.5) gives

$$\beta_0^{(1)}(x_2, \cdots, x_n) = \beta$$

almost everywhere, while the functional form of $g(x_1, \cdots, x_n)$ above implies

$$\begin{aligned} \beta_0^{(1)}(x_2, \cdots, x_n) &= \beta, & &\max_{j \neq 1} x_j \leqq N_\beta, \\ &= 1, & &\max_{j \neq 1} x_j > N_\beta. \end{aligned}$$

These are obviously in conflict unless $n = 1$ or $\beta = 0, 1$. This completes the proof of Theorem 1.

3.3. *Examples of normal confidence bounds*. Although an exact β-level confidence bound satisfying assumptions 2.1 and 2.2 does not exist, bounds with at least β confidence do exist; an example of one was obtained in the course of the proof of Theorem 1, namely,

$$g(x_1, \cdots, x_n) = \max x_i + N_{1-\beta}\, \sigma.$$

It is easily seen from the form of $\beta_0^{(1)}(x_2, \cdots, x_n)$ that this bound has at least β confidence,

$$\begin{aligned} \beta_0^{(1)}(x_2, \cdots, x_n) &= \beta, & \max_{j \neq i} x_j \leqq N_\beta, \\ &= 1, & \max_{j \neq i} x_j > N_\beta. \end{aligned}$$

The confidence level is

$$E\{\beta_0^{(1)}(X_2, \cdots, X_n) \mid \max_{j \neq 1} \mu_j \leqq 0\} \geqq \beta.$$

We define a bound g having confidence at least β to be uniformly better than a bound g' having confidence at least β, if $g \leqq g'$ for all $(x_1, \cdots, x_n)$, and $g < g'$ on a set of positive measure. It is not difficult to see that bounds, uniformly better than the example above, do not exist. This obtains from the following simple property of the normal distribution. Let Y be normal with mean θ and variance 1; then for δ positive, all the probability less than C can be made arbitrarily small with respect to the probability in any small neighborhood of $c + \delta$ by taking θ large enough.

Since it may be desirable to obtain bounds other than the example given above, we outline the procedure. For spherically symmetric normal distributions in R^n having variance σ^2 and mean $(\mu_1, \cdots, \mu_n)$ with $\max \mu_i = 0$, we look for a region whose size is greater than or equal to β and whose characteristic function $\phi(x_1 + \delta, \cdots, x_n + \delta)$ is monotone nondecreasing in δ; then a β level bound $g(x_1, \cdots, x_n)$ satisfying Assumption 2.1* is the following:

$$g(x_1, \cdots, x_n) = \delta'$$

where δ' is the value of δ at which $\phi_0(x_1 - \delta, \cdots, x_n - \delta)$ jumps from 0 to 1.

3.4. *Bounds for nonnormal distributions.* As remarked in Section 2, confidence bounds for $\max \mu_i$ may be wanted for distributions other than the normal; $\mu_1, \cdots, \mu_n$ would of course be values of the location parameter corresponding to the random variables $X_1, \cdots, X_n$. Consider the density function $f(x - \mu)$; we shall say it is boundedly complete (one-sided) if

$$\int_{-\infty}^{\infty} g(x) f(x - \mu)\, dx = 0$$

for any dense set of $\mu < 0$ and $|g(x)| < M$ imply $g(x) = 0$ almost everywhere. From a theorem of Lehmann and Scheffé which was mentioned in [1], we can conclude that if $f(x - \mu)$ is boundedly complete (one-sided) then

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \cdots, x_n) \prod_i f(x_i - \mu_i) \prod_i dx_i = 0$$

for all $\mu_1, \cdots, \mu_n < 0$ and $|g(x_1, \cdots, x_n)| < M$ imply $g(x_1, \cdots, x_n) = 0$ almost everywhere. This conclusion takes the place of Lemma 1 for the following theorem:

THEOREM 2. *If $X_1, \cdots, X_n$ are independent and have probability density functions $f(x - \mu_1), \cdots, f(x - \mu_n)$, where $f(x - \mu)$ is bounded complete (one-sided), then there does not exist (unless $\beta = 0, 1$, or $n = 1$) a measurable function $g(x_1, \cdots, x_n)$, satisfying Assumptions 2.1 and 2.2, which is an exact β confidence bound for max μ_i.*

PROOF: The proof is essentially that of Theorem 1.

4. Introduction to Problem B. The second problem is to find a confidence interval for a set of means; if $X_1, \cdots, X_n$ are normally and independently distributed with known variance σ^2 and unknown means $\mu_1, \cdots, \mu_n$, Problem B is to find two functions $g(x_1, \cdots, x_n)$, $h(x_1, \cdots, x_n)$ such that

$$\Pr\{g(X_1, \cdots, X_n) \geqq \mu_1, \cdots, \mu_n \geqq h(X_1, \cdots, X_n)\} \geqq \beta.$$

We also study the problem of finding an exact β-level confidence interval for which the above condition is replaced by

$$\Pr\{g(X_1, \cdots, X_n) \geqq \mu_1, \cdots, \mu_n \geqq h(X_1, \cdots, X_n)\} = \beta.$$

In Section 5.3 we establish the nonexistence of exact β-level confidence intervals among pairs of functions (h, g) satisfying several moderate and reasonable restrictions; these restrictions are:

ASSUMPTION 4.1. *The functions $g(x_1, \cdots, x_n)$ and $h(x_1, \cdots, x_n)$ satisfy the equations*

$$g(x_r + \delta, \cdots, x_n + \delta) = g(x_1, \cdots, x_n) + \delta,$$

$$h(x_1 + \delta, \cdots, x_n + \delta = h(x_1, \cdots, x_n) + \delta$$

for all $x_1, \cdots, x_n, \delta$.

ASSUMPTION 4.2. *The equation*

$$g(x_1, \cdots, x_n) = -h(-x_1, \cdots, -x_n)$$

holds for all $x_1, \cdots, x_n$.

ASSUMPTION 4.3. *The functions $g(x_1, \cdots, x_n)$ and $h(x_1, \cdots, x_n)$ are symmetric functions.*

ASSUMPTION 4.4. *If $x_j = \max x_i$, then the function $g(x_1, \cdots, x_n)$ satisfies*

$$g(x_1, \cdots, x_n) \leqq g(x_1, \cdots, x_{j-1}, x_j + \delta, x_{j+1}, \cdots, x_n)$$

for any positive δ.

ASSUMPTION 4.5. *For all $x_1, \cdots, x_n$, $g(x_1, \cdots, x_n) \geqq \bar{x} + \epsilon_g$, where $\bar{x} = \sum x_i/n$ and $\epsilon_g > 0$, may depend on g but not on $x_1, \cdots, x_n$.*

As a corollary to Theorem 3 we obtain a confidence interval for the means which has at least β confidence; it is (4.1) $(h, g) = (\min x_i - N_{\frac{1}{2}(1-\beta)}\,\sigma, \max x_i + N_{\frac{1}{2}(1-\beta)}\,\sigma)$ where N_α is the α point of the unit normal. Also in section 5 we indicate the procedure for constructing intervals having at least β confidence.

5. Analysis of Problem B.

5.1. *Justification of assumptions.* The first three assumptions (4.1, 4.2 and 4.3) are obtained by applying the principle of cogredience to the problem. The set of transformations

$$x_i' = x_i + C, i = 1, \cdots, n, \qquad C \varepsilon R^1$$

produces the conditions contained in Assumption 4.1. Similarly the transformations

$$x_i' = -x_i, i = 1, \cdots, n,$$

$$x_i' = x_{j_i}, i = 1, \cdots, n,$$

for all permutations $(j_1, \cdots, j_n)$ of $(1, \cdots, n)$ produce respectively the conditions of Assumptions 4.2 and 4.3.

Assumption 4.4 is similar in form and justification to Assumption 2.2. Assumption 4.5 is not too restrictive for practical confidence intervals: it is introduced merely from necessity in the proof. Nevertheless, it seems reasonable to suppose that Assumption 4.5 is not essential for the conclusions of the theorem.

5.2. *Characteristic functions.* A characteristic function similar to (3.1) could be defined for the interval (h, g). However, the symmetry introduced by Assumption 4.2 enables us to use the characteristic function (3.1); for $g(x_1, \cdots, x_n)$ in (h, g) we define $\phi_\theta(x_1, \cdots, x_n)$ as in (3.1).

The present assumptions yield for $\phi_\theta(x_1, \cdots, x_n)$ the properties derived in Section 3.1, namely,

(1) $\phi_\theta(x_1 + \delta, \cdots, x_n + \delta)$ is monotone nondecreasing as a function of δ, and

(2) for points $(x_1, \cdots, x_n) \varepsilon S_i$, $\phi_\theta(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ is a monotone nondecreasing function of x_i.

5.3. *Theorem for normal distributions.* To establish the nonexistence of exact β-level confidence intervals satisfying Assumptions 4.1 to 4.5, we have

THEOREM 3. *If $X_1, \cdots, X_n$ are normally and independently distributed with means $\mu_1, \cdots, \mu_n$ and variance σ^2, there does not exist (unless $\beta = 0, 1$ or $n = 1$) a pair of measurable functions (g, h) which satisfies Assumptions* 4.1, 4.2, 4.3, 4.4, 4.5 *and which is an exact β-level confidence interval for the set $\{\mu_1, \cdots, \mu_n\}$.*

PROOF: The proof is somewhat different from that used in Theorem 1, but several results obtained in the course of that proof are used here.

Let $\sigma^2 = 1$ without loss of generality. We consider a pair of functions (h, g) satisfying Assumptions 4.1 to 4.5, and, assuming (h, g) is an exact β level confidence interval, we shall find that a contradiction results.

For the characteristic function $\phi_\theta(x_1, \cdots, x_n)$ define according to (3.4) a conditional expectation $\beta_\theta^{(i)}(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$. In the following expressions we shall use a symmetric multivariate normal distribution with variance 1 and mean given after the condition bars. Using Assumption 4.1, we have

$$\beta = \Pr\{g(X_1, \cdots, X_n) \geqq 0, h(X_1 + \theta_{n-1}, \cdots, X_n + \theta_{n-1}) \leqq 0 \mid (\theta, -\theta_1, \cdots, -\theta_{n-1})\}$$

if we let $0 \leqq \theta_1 \leqq \cdots \leqq \theta_{n-1}$. Using Assumptions 4.1, 4.2, 4.3,

$$\begin{aligned}
1 - \beta &= E\{(1 - \phi_0(X_1, \cdots, X_n)) \mid (0, -\theta_1, \cdots, -\theta_{n-1})\} \\
&\quad + E\{(1 - \phi_0(X_1 - \theta_{n-1}, \cdots, X_n - \theta_{n-1})) \mid (0, \theta_1, \cdots, \theta_{n-1})\} \\
&\quad - E\{(1 - \phi_0(X_1, \cdots, X_n))(1 - \phi_0(-X_1 - \theta_{n-1}, \cdots, -X_n - \theta_{n-1})) \\
&\qquad\qquad \mid (0, -\theta_1, \cdots, -\theta_{n-1})\} \\
&= E\{(1 - \phi_0(X_1, \cdots, X_n) \mid (0, -\theta_1, \cdots, -\theta_{n-1})\} \\
&\quad + E\{(1 - \phi_0(X_1, \cdots, X_n) \\
&\qquad\qquad \mid (-\theta_{n-1}, -(\theta_{n-1} - \theta_1), \cdots, -(\theta_{n-1} - \theta_{n-2}), 0)\} \\
&\quad - E\{(1 - \phi_0(X_1, \cdots, X_n))(1 - \phi_0(-X_1 - \theta_{n-1}, \cdots, -X_n - \theta_{n-1})) \\
&\qquad\qquad \mid (0, -\theta_1, \cdots, -\theta_{n-1})\}.
\end{aligned}$$

If we restrict the values of the θ's it is possible to make the third term on the right hand side of the equation equal to zero. From Assumption 4.5 the first factor $1 - \phi_0(x_1, \cdots, x_n)$ is equal to zero if $g(x_1, \cdots, x_n) \geqq 0$ or if $\bar{x} + \epsilon_g \geqq 0$. Similarly the second factor $1 - \phi_0(-x_1 - \theta_{n-1}, \cdots, -x_n - \theta_{n-1})$ is equal to zero, if $-x_i - \theta_{n-1} + \epsilon_g \geqq 0$ or if $\bar{x} + \epsilon_g \leqq 2\epsilon_g - \theta_{n-1}$. The product of the two factor will certainly be zero if $\theta_{n-1} < 2\,\epsilon_g$. Therefore we have

$$\begin{aligned}
1 - \beta &= E\{(1 - \phi_0(x_1, \cdots, x_n)) \mid (0, -\theta_1, \cdots, -\theta_{n-1})\} \\
&\quad + E\{(1 - \phi_0(x_1, \cdots, x_n)) \mid (0, -(\theta_{n-1} - \theta_{n-2}), \cdots, -(\theta_{n-1} - \theta_1))\}
\end{aligned}$$

for all $\theta_1, \cdots, \theta_{n-1}$ satisfying $0 \leqq \theta_1 \leqq \cdots \leqq \theta_{n-1} < 2\epsilon_g$.

We now derive a property of the conditional expectation $\beta_0^{(1)}(x_2, \cdots, x_n) = \beta(x_2, \cdots, x_n)$. We have

$$\begin{aligned}
1 - \beta &= \frac{1}{(2\pi)^{\frac{1}{2}(n-1)}} \int (1 - \beta(x_2, \cdots, x_n)) \exp\left[-\frac{1}{2}\left\{\sum_2^n (x_i + \theta_{i-1})^2\right\}\right] dx_2 \cdots dx_n \\
&\quad + \frac{1}{(2\pi)^{\frac{1}{2}(n-1)}} \int (1 - \beta(x_2, \cdots, x_n)) \\
&\quad \cdot \exp\left[-\frac{1}{2}\left\{\sum_2^{n-1} (x_i + \theta_{n-1} - \theta_{i-1})^2 + (x_n + \theta_{n-1})^2\right\}\right] dx_2 \cdots dx_n.
\end{aligned}$$

We note that the following functions satisfy the conditions for a pdf in R^{n-1}:

$$\begin{aligned}
f_1 &= \frac{1}{2(2\pi)^{\frac{1}{2}(n-1)}} \left\{\exp\left[-\frac{1}{2}\sum_2^n (x_i + \theta_{i-1})^2\right]\right. \\
&\qquad \left. + \exp \int \left[-\frac{1}{2}\sum_2^{n-1} (x_i + \theta_{n-1} - \theta_{i-1})^2 - \tfrac{1}{2}(x_n + \theta_n - 1)^2\right]\right\},
\end{aligned}$$

$$f_2 = 2\,\frac{1 - \beta(x_2, \cdots, x_n)}{1 - \beta} f_1.$$

For the integral

$$\int_{R^{n-1}} f_i \, dx_2 \cdots dx_n = 1,$$

the conditions are satisfied for differentiating any number of times under the sign of integration with respect to $\theta_1, \cdots, \theta_{n-1}$. If we set $\theta_1 = \cdots = \theta_{n-1} = 0$ in the equation

$$\int \frac{\partial^{r_2+\cdots+r_n}}{\partial\theta_1^{r_2} \cdots \partial\theta_{n-1}^{r_n}} f_i \, dx_2 \cdots dx_n = 0,$$

we obtain equations from which all the moments of f_i (with all θ's equal zero) can be obtained. However, the equations do not depend on i; hence f_1 and f_2 have identical moments. But for these multivariate normal moments, the density corresponding to them is unique (generalization of moment condition on p. 176 in [2]); therefore,

$$1 - \beta(x_2, \cdots, x_n) = \tfrac{1}{2}(1 - \beta)$$

or

$$\beta(x_2, \cdots, x_n) = \tfrac{1}{2}(1 + \beta) = \beta^*.$$

Now if we use part of the proof of Theorem 1 from formula (3.5) to formula (3.12), we obtain

$$\begin{aligned} g(x_1, \cdots, x_n) &= \max x_i + N_{1-\beta^*}, \\ &= \max x_i + N_{\frac{1}{2}(1-\beta)}, \end{aligned}$$

and by Assumption 4.2, we have

$$h(x_1, \cdots, x_n) = \min x_i - N_{\frac{1}{2}(1-\beta)}.$$

Therefore

$$(h,g) = (\min x_i - N_{\frac{1}{2}(1-\beta)}, \max x_i + N_{\frac{1}{2}(1-\beta)}).$$

It is easily seen that for this interval the confidence level is greater than β (unless $\beta = 0$, 1, or $n = 1$). Since this is a contradiction the theorem is proved.

5.4. *Example of normal confidence intervals.* Intervals having *at least* β confidence do exist; for example

$$(\min x_i - \sigma N_{\frac{1}{2}(1-\beta)}, \max x_i + \sigma N_{\frac{1}{2}(1-\beta)}).$$

The confidence level for this interval is always larger than β and it seems reasonable to expect that it is bounded away from β. In other words the above interval might be refined by using a constant smaller than $N_{\frac{1}{2}(1-\beta)}$. The answer to this question will most likely be obtained only by applying a numerical procedure analogous to that described at the end of Section 3.4.

Any bounds for Problem A can be used to provide an interval for the present problem. Let $1 - \beta = \alpha_1 + \alpha_2$ where α_1 and α_2 are positive, and let $g_1(x_1, \cdots, x_n)$, $G_2(x_1, \cdots, x_n)$ be at least $1 - \alpha_1$, $1 - \alpha_2$ confidence bounds for problem A. Then an interval having at least β confidence is

$$(-g_2(-x_1, \cdots, -x_n), g_1(x_1, \cdots, x_n))$$

This follows from the argument at the beginning of the proof of Theorem 3.

REFERENCES

[1] E. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiased estimation," *Sankhyā*, Vol. 10 (1950), pp. 305-340.
[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

SEQUENTIAL MINIMAX ESTIMATION FOR THE RECTANGULAR DISTRIBUTION WITH UNKNOWN RANGE[1]

By J. Kiefer

Cornell University

1. Summary. This paper is concerned with sequential minimax estimation of the parameter $\theta (0 < \theta < \infty)$ of the density function (3.1) when the observations are independently and identically distributed with this density, each observation costs the same amount $c > 0$, and the weight function is as given in Section 2. A procedure requiring a fixed sample size is shown to be a minimax solution for this problem.

2. Introduction. An important problem in the theory of statistical decision functions[2] is that of minimax sequential estimation of the parameter of an (unknown) member of a given family of distribution functions when the observations are taken on chance variables which are independently and identically distributed and when the cost of taking n observations is cn (with $c > 0$) regardless of the way in which they are taken. This problem was solved for the case of point estimation of the mean of the rectangular distribution from $\theta - \frac{1}{2}$ to $\theta + \frac{1}{2}$ $(-\infty < \theta < \infty)$, for weight function $W(\theta, d) = (\theta - d)^2$ by Wald [1]; the minimax sequential estimation problem for the normal distribution was solved for a variety of terminal decision spaces and weight functions by Wolfowitz [2] (see also [3]); certain extensions and modifications of the results of both of these cases were given by Blyth [4].

The present paper is devoted to a problem of sequential minimax estimation for the case where the family of possible distribution functions consists of all distributions for which the successive observations are independently and identically distributed with rectangular density function from 0 to θ (equation (3.1)) for $\theta \varepsilon \Omega = \{\theta \mid 0 < \theta < \infty\}$ and where the cost of taking n observations is $cn (c > 0)$ regardless of the way in which the observations are taken. The object is to estimate θ, the terminal decision space being $D = \{d \mid 0 \leqq d < \infty\}$. The weight function is $W(\theta, d) = [(\theta - d)/\theta]^2$; i.e., the loss incurred by making decision d when θ is the true parameter is the square of the fractional error in estimating θ. Thus, the minimax problem considered in this paper is that of finding a sequential estimation procedure which minimizes $\sup_\theta \{cE_\theta(n) + E_\theta[(\theta - d)/\theta]^2\}$. A word is in order concerning our choice of weight function. The reason we do not study the problem for such weight functions as $|\theta - d|$, $(\theta - d)^2$, or $[(\theta - d)^2/\theta]$ is that for such weight functions the supremum of the risk over all $\theta \varepsilon \Omega$ is infinite for every decision function, so that

[1] Research under a contract with the Office of Naval Research.

[2] See Wald [1] for an exposition of this theory and an explanation of the nomenclature used herein.

every decision function is minimax. In addition, weight functions which depend only on d/θ (such as $[(\theta - d)/\theta]^2$) have a structure which essentially simplifies matters when estimating a scale parameter. On the other hand, it does not seem convenient in the present case to consider simultaneously a large class of weight functions as was possible in the cases of symmetrical densities studied in [2] and [4]. We therefore treat only one typical weight function here, noting that the same method should be applicable to many others.

With Ω, D, and $W(\theta, d)$ as described above, we shall prove that there is a minimax solution for which a fixed number of observations is taken. Specifically, the function $r(m)$ of (3.20) (which is the constant risk corresponding to taking a sample of fixed sample size m and then estimating θ by the expression of (2.1) with m for m_0) has at most two minima (if there are two, they are for successive values of m; moreover, there is only one minimum for all but a denumerable set of values of c). A minimax decision function is given by taking m_0 observations $y_1, y_2, \cdots y_{m_0}$, where $r(m_0)$ is the minimum of $r(m)$ (if there are two minima, at m_0 and $m_0 + 1$, one may randomize in any way between the decisions to take m_0 or $m_0 + 1$ observations); and by then estimating θ by

$$\text{(2.1)} \qquad \frac{m_0 + 2}{m_0 + 1} \max (y_1, \cdots, y_{m_0})$$

if $m_0 > 0$ (we replace m_0 by $m_0 + 1$ throughout (2.1) if the latter number of observations is taken when there are two minima), and by 0 if $m_0 = 0$. The risk corresponding to this decision function is then $r(m_0)$ for all values of $\theta \, \varepsilon \, \Omega$. It follows, incidentally, that this decision function is uniformly best among all cogredient procedures (see [4]). It is also a minimax solution for some related problems discussed in Section 3 of [4].

The method of proof is to calculate a lower bound on the Bayes risk when the a priori density on Ω is given by (3.4). It follows from (3.24) that as the parameter a of (3.4) approaches zero, the corresponding Bayes risk approaches $r(m_0)$; hence, by an argument like that of [1], p. 167, the procedure described in the previous paragraph is a minimax solution. The lower bound (3.24) is calculated in detail, since the necessary steps in its calculation differ somewhat from those of [1], [2], and [4]. We also note that, in this case of estimating a scale parameter, the tool used in [1], [2], and [4] of attempting to attain a "uniform a priori distribution on the real line" in the location parameter case is replaced by trying to attain the "a priori density" $1/\theta$. The proof is somewhat shortened by restricting the positive range of $\lambda_a(\theta)$ to values $\theta < 1$. This asymmetry manifests itself in the fact that the estimator of (3.7) does not tend to a minimax solution as $a \to 0$.

The fact that $\lambda_a(\theta)$ is positive only for $\theta < 1$ also shows that the fixed sample procedure described above is minimax for the problem of estimating θ when the above setup is altered by making $\Omega = \{\theta \mid 0 < \theta < b\}$, where $0 < b < \infty$: the argument of Section 3 shows this for $b = 1$, and the result for general $b = b'$ follows immediately from the case $b = 1$ if one considers there the problem of estimating $b'\theta$ from the sequence $\{b'Y_i\}$ of chance variables. Similarly, by considering for each value of a in Section 3 the problem of estimating $ba\theta$ from the sequence $\{baY_i\}$, one sees that our fixed sample procedure is also minimax for the problem of estimating θ when our original setup is altered by making $\Omega = \{\theta \mid b < \theta < \infty\}$. However, the given procedure is obviously not admissible if $m_0 > 0$ (or $m_0 \geqq 0$ in the second case): for example, a trivially better procedure in the first case when $m_0 > 0$ is to estimate θ by b whenever the expression of (2.1) is $> b$.

Finally, we remark that the problem of estimating θ for the case where the $f(y; \theta)$ of (3.1) is replaced by $1/(2\theta)$ for $-\theta < y < \theta$, is obviously identical to the one we consider: one has only to note that after n observations a sufficient statistic is still given by (3.2) if only Y_i is replaced by $|Y_i|$ for $i = 1, \cdots, n$. It is also of interest to note that our problem may be translated (by considering $T_i = e^{-Y_i}$, $\phi = e^{-\theta}$) into that of sequential minimax estimation of the parameter ϕ of the density $e^{-(t-\phi)}$ for $t > \phi$, 0 otherwise $(-\infty < \phi < \infty)$, when the weight function is $W(\phi, d) = (1 - e^{-(d-\phi)})^2$.

3. Calculations. For brevity, we shall throughout this section state the values of density functions and discrete probability functions only over the domains where they are positive. Let $Y_1, Y_2, \cdots$ be a sequence of independently and identically distributed chance variables, each with density function

$$f(y; \theta) = 1/\theta \qquad 0 < y < \theta, \tag{3.1}$$

where $\theta \,\varepsilon\, \Omega = \{\theta \mid 0 < \theta < \infty\}$. Define

$$X_n = \max\{Y_1, \cdots, Y_n\}. \tag{3.2}$$

Clearly, if observations $y_1, \cdots, y_n$ on $Y_1, \cdots, Y_n$ are taken, then X_n is a sufficient statistic for θ; i.e., for any a priori probability distribution on Ω, the a posteriori distribution of θ depends on $y_1, \cdots, y_n$ only through the value x_n taken on by X_n. Thus, in constructing sequential Bayes solutions, we may restrict ourselves to decision functions for which the (perhaps randomized) rule for stopping and estimation depends, after n observations, only on x_n. The density function of X_n is given by

$$g_n(x; \theta) = \frac{nx^{n-1}}{\theta^n}, \qquad 0 < x < \theta. \tag{3.3}$$

For $0 < a < 1$, we define

$$\lambda_a(\theta) = \frac{1}{\log(1/a)} \frac{1}{\theta}, \qquad a < \theta < 1. \tag{3.4}$$

If $\lambda_a(\theta)$ is the a priori density function on Ω and $y_1, \cdots, y_n$ have been observed, the a posteriori density of θ given that $X_n = x$ is easily computed to be

$$h_n(\theta \mid X_n = x) = \frac{nz^n}{1 - z^n} \cdot \frac{1}{\theta^{n+1}}, \qquad z < \theta < 1, \tag{3.5}$$

where $z = \max(a, x)$ and we note that $P\{z < 1\} = 1$.

The a posteriori loss (excluding cost of experimentation) if one stops after n observations and uses d to estimate θ, is

$$
\begin{aligned}
W_n^*(d, z) &= \int_0^1 \left(\frac{d-\theta}{\theta}\right)^2 h_n(\theta \mid X_n = x)\, d\theta \\
&= 1 + \frac{n}{z^2(1-z^n)}\left[z^{n+2}\left(\frac{2d}{n+1} - \frac{d^2}{n+2}\right) - \left(\frac{2dz}{n+1} - \frac{d^2}{n+2}\right)\right].
\end{aligned} \tag{3.6}
$$

The unique minimum of W_n^* with respect to d is easily seen to occur for

$$d = \frac{n+2}{n+1} \cdot \frac{1-z^{n+1}}{1-z^{n+2}} \cdot z, \tag{3.7}$$

the corresponding value of W_n^* being

$$W_n^{**}(z) = 1 - \frac{n(n+2)}{(n+1)^2} \cdot \frac{(1-z^{n+1})^2}{(1-z^n)(1-z^{n+2})}. \tag{3.8}$$

For $n = 0$, the integral in (3.6) must be altered by replacing h_n by λ_a; the final expression must be changed accordingly. Equation (3.7) then holds with $z = a$, and (3.8) becomes $1 - 2(1-a)/[(1+a)\log(1/a)]$.

Next we note that when $f(y;\theta)$ is the density of each Y_i, the conditional distribution function of X_n given that $X_{n-1} = u$ assigns probability mass u/θ at the point $x = u$ and density $1/\theta$ for $u < x < \theta$. For $n = 1$, the distribution of X_1 is of course given by the density $f(x;\theta)$. We conclude that if $\lambda_a(\theta)$ is the a priori density on Ω, the distribution of X_1 is given by the density

$$p_1(x) = \int_0^1 f(x;\theta)\lambda_a(\theta)\, d\theta = \begin{cases} \dfrac{1}{\log(1/a)} \cdot \dfrac{(1-a)}{a} & \text{if } x \leqq a, \\[2ex] \dfrac{1}{\log(1/a)} \cdot \dfrac{(1-x)}{x} & \text{if } a < x < 1; \end{cases} \tag{3.9}$$

and that (using (3.5) with n replaced by $n-1$), for $n > 1$, the conditional distribution of X_n given that $\lambda_a(\theta)$ is the a priori density and $X_{n-1} = u$, is given, if $u \leqq a$, by

$$
\begin{aligned}
P_n\{X = u\} &= \frac{n-1}{n} \cdot \frac{1-a^n}{1-a^{n-1}} \cdot \frac{u}{a}, \\
p_n(x \mid u) &= \begin{cases} \dfrac{n-1}{n} \cdot \dfrac{1-a^n}{1-a^{n-1}} \cdot \dfrac{1}{a}, & u < x \leqq a, \\[2ex] \dfrac{n-1}{n} \cdot \dfrac{1-x^n}{1-a^{n-1}} \cdot \dfrac{a^{n-1}}{x^n}, & a < x < 1; \end{cases}
\end{aligned} \tag{3.10}
$$

and, if $u > a$, by

$$
\begin{aligned}
P_n\{X = u\} &= \frac{n-1}{n} \cdot \frac{1-u^n}{1-u^{n-1}}, \\
p_n(x \mid u) &= \frac{n-1}{n} \cdot \frac{1-x^n}{1-u^{n-1}} \cdot \frac{u^{n-1}}{x^n}, \qquad u < x < 1;
\end{aligned} \tag{3.11}
$$

where in each case P_n is the probability mass at $x = u$ and $p_n(x \mid u)$ is the density elsewhere.

Equations (3.10) and (3.11) yield for the conditional distribution of $Z_n = \max(X_n, a)$ given that $\lambda_a(\theta)$ is the a priori density and that $Z_{n-1} = v$, for all $n > 1$,

$$
\begin{aligned}
&Q_n\{Z = v\} = \frac{n-1}{n} \cdot \frac{1 - v^n}{1 - v^{n-1}}, \\
(3.12)\qquad &q_n(z \mid v) = \frac{n-1}{n} \cdot \frac{1 - z^n}{1 - v^{n-1}} \cdot \frac{v^{n-1}}{z^n}, \qquad v < z < 1,
\end{aligned}
$$

where again q_n is a density and Q_n is the probability mass at $z = v$.

Let $\overline{W}_{n-1}(v)$ be the conditional expected value of $W_n^{**}(Z_n)$ given that $\lambda_a(\theta)$ is the a priori density and that $Z_{n-1} = v$ (where we define $Z_0 = a$). Using (3.8) and (3.9), we have

$$
\begin{aligned}
\overline{W}_0(a) &= E\{W_1^{**}(Z_1)\} \\
&= W_1^{**}(a) \int_0^a p_1(z)\, dz + \int_a^1 W_1^{**}(z) p_1(z)\, dz \\
(3.13a)\qquad &= 1 - \frac{3}{4 \log (1/a)} \left\{ \frac{(1 - a^2)^2}{(1 - a^3)} + \int_a^1 \frac{(1 - z^2)^2}{z(1 - z^3)}\, dz \right\} \\
&< 1 - \frac{3}{4 \log (1/a)} \int_a^1 \left[\frac{1}{z} - 1 + \frac{(1 - z)^2}{(1 - z^3)} \right] dz \\
&< 1 - \frac{3}{4 \log (1/a)} \left[\log \frac{1}{a} - (1 - a) \right] < \tfrac{1}{4} + \frac{1}{\log (1/a)}.
\end{aligned}
$$

For $n > 1$, we have from (3.8) and (3.12),

$$
\begin{aligned}
\overline{W}_{n-1}(v) &= E\{W_n^{**}(Z_n) \mid v\} \\
&= W_n^{**}(v) Q_n(Z = v) + \int_v^1 W_n^{**}(z) q_n(z \mid v)\, dz \\
(3.13b)\qquad &= 1 - \frac{(n-1)(n+2)}{(n+1)^2(1 - v^{n-1})} \\
&\qquad \cdot \left\{ \frac{(1 - v^{n+1})^2}{(1 - v^{n+2})} + v^{n-1} \int_v^1 \frac{(1 - z^{n+1})^2}{z^n(1 - z^{n+2})}\, dz \right\}.
\end{aligned}
$$

The term in the last set of braces in (3.13b) may be written as

$$
\begin{aligned}
&1 - v^{n-1} + \frac{v^{n-1}(1 - v)[1 + v - v^2 - v^{n+2}]}{(1 - v^{n+2})} \\
&\qquad + v^{n-1} \int_v^1 \left[\frac{1}{z^n} - \frac{z(2 - z - z^{n+1})}{(1 - z^{n+2})} \right] dz \\
(3.14)\qquad & \\
&> (1 - v^{n-1}) + \frac{1}{n-1}(1 - v^{n-1}) - v^{n-1} \int_v^1 2z\, dz \\
&= \frac{n}{n-1}(1 - v^{n-1}) - v^{n-1}(1 - v^2).
\end{aligned}
$$

We conclude that, whenever $n > 1$,

$$\overline{W}_{n-1}(v) < 1 - \frac{n(n+2)}{(n+1)^2} + \frac{(n-1)(n+2)}{(n+1)^2} \cdot \frac{v^{n-1}(1-v^2)}{(1-v)^{n-1}} \tag{3.15}$$
$$< \frac{1}{(n+1)^2} + \frac{1}{\log(1/v)},$$

where in the last step we have used the fact that $(n-1)(n+2)(n+1)^{-2} < \frac{1}{2}$ if $n = 2$ and < 1 otherwise, that $(1-v^2)(1-v^{n-1})^{-1} < 2$ if $n = 2$ and $\leqq 1$ otherwise, and that if $n > 1$ we have $v^{n-1} \leqq v < (\log 1/v)^{-1}$. From (3.13a) and (3.15), we have for all $n > 0$,

$$\overline{W}_{n-1}(v) < \frac{1}{(n+1)^2} + \frac{1}{\log(1/v)}. \tag{3.16}$$

Similarly, we have from (3.8) for $n > 1$,

$$W^{**}_{n-1}(v) = 1 - \frac{(n-1)(n+1)}{n^2} \cdot \frac{(1-v^n)^2}{(1-v^{n-1})(1-v^{n+1})}$$
$$= 1 - \frac{(n-1)(n+1)}{n^2}\left[1 + \frac{v^{n-1}(1-v)^2}{(1-v^{n-1})(1-N^{n+1})}\right] \tag{3.17}$$
$$> \frac{1}{n^2} - v^{n-1} \geqq \frac{1}{n^2} - v > \frac{1}{n^2} - \frac{1}{\log(1/v)};$$

and, for $n = 1$ (putting $v = a$),

$$W^{**}_0(v) = 1 - \frac{2(1-a)}{(1+a)\log(1/a)} > 1 - \frac{2}{\log(1/v)}. \tag{3.18}$$

Combining (3.16), (3.17), and (3.18), we have for all $m \geqq 0$,

$$W^{**}_m(v) - \overline{W}_m(v) > \frac{2m+3}{(m+1)^2(m+2)^2} - \frac{3}{\log(1/v)}. \tag{3.19}$$

We now define, for all integers $m \geqq 0$,

$$r(m) = cm + \frac{1}{(m+1)^2}. \tag{3.20}$$

We note that $r(m+1) - r(m) = c - (2m+3)/((m+1)^2(m+2)^2)$. The function $r(m)$ evidently has at most two minima (if there are two, they are for consecutive values of m). Denote by m_0 the first integer for which $r(m_0)$ is a minimum. Let ϵ $(0 < \epsilon < 1)$ be such that $3\epsilon < r(m_0 - 1) - r(m_0)$ (if $m_0 = 0$, the last restriction is omitted). Let $d = e^{-1/\epsilon}$ and $a = e^{-1/\epsilon^2}$. Let m_1 be the smallest integer not less than $1/c$.

For any integer $K > 0$, if $\lambda_a(\theta)$ is the a priori density we have (noting that $d > a$)

$$P\{X_K \geqq d\} = \int_d^1 \int_x^1 g_K(x; \theta)\lambda_a(\theta)\, d\theta\, dx = \int_d^1 \frac{1 - x^K}{x \log(1/a)}\, dx \tag{3.21}$$

$$= \frac{\log(1/d)}{\log(1/a)} - \frac{1 - d^K}{K \log(1/a)} < \epsilon.$$

We note that, after m observations ($m = 0, 1, \cdots$, ad inf. and putting $v = a$ if $m = 0$), any Bayes solution will certainly prescribe taking another observation if $W_m^{**}(v) - \overline{W}_m(v) - c > 0$, since this quantity is the a posteriori expected saving over stopping after m observations if instead one takes one additional observation and then stops and makes the best terminal decision.

We also note that, since $(\log 1/a)^{-1} = \epsilon^2 < \epsilon$, it follows from (3.21) that, when $\lambda_a(\theta)$ is the a priori density,

$$P\left\{\frac{1}{\log(1/Z_i)} < \epsilon \text{ for } i = 1, 2, \cdots, m_0 + m_1\right\} = P\left\{\frac{1}{\log(1/Z_{m_0+m_1})} < \epsilon\right\}$$

$$= P\left\{\frac{1}{\log(1/X_{m_0+m_1})} < \epsilon\right\} = P\{X_{m_0+m_1} < d\} > 1 - \epsilon.$$

Since $r(m-1) - r(m)$ is a decreasing function of $m (m > 0)$ and since $3\epsilon < r(m_0 - 1) - r(m_0)$, we conclude that, if $m_0 > 0$, the event

$$\frac{1}{\log(1/Z_{m_0+m_1})} < \epsilon \tag{3.23}$$

entails the event $(\log(1/Z_{m_0-1}))^{-1} < \epsilon$, which entails $3(\log(1/Z_{m_0-1}))^{-1} < r(m_0 - 1) - r(m_0)$; or, equivalently, $-3(\log(1/Z_i))^{-1} + r(i) - r(i+1) > 0$ for $i = 0, 1, \cdots, m_0 - 1$. Finally, it follows from (3.19) that this entails the event $W_i^{**}(v) - \overline{W}_i(v) - c > 0$ for $i = 0, 1, \cdots, m_0 - 1$; and, for any Bayes solution relative to $\lambda_a(\theta)$, this entails the event that at least m_0 observations will be taken. Furthermore, the last statement is always true for $m_0 = 0$.

Similarly, we note from (3.17) and (3.18) that the event (3.23) certainly entails the event $W_i^*(v) > (1/(1+i)^2) - 2\epsilon$ for $i = m_0, m_0 + 1, \cdots, m_0 + m_1$. That is, if a terminal decision is made after exactly i observations ($i = m_0, \cdots, m_0 + m_1$), the total a posteriori loss plus cost of experimentation will be $> ci + (1/(1+i)^2) - 2\epsilon \geqq cm_0 + (1/(1+m_0)^2) - 2\epsilon$. Moreover, it follows from the definition of m_1 that this last expression is less than the cost of experimentation alone if more than $m_0 + m_1$ observations are taken.

To summarize, then, the event (3.23) implies for any Bayes solution relative to $\lambda_a(\theta)$ that the experiment will terminate with a total a posteriori loss plus cost of experimentation exceeding $cm_0 + (1/(1+m_0)^2) - 2\epsilon$. But it follows from (3.22) that (3.23) occurs with probability $>1 - \epsilon$. Since $m_0 c + (1/(m_0+1)^2) \leqq 1$, it follows that the Bayes risk relative to $\lambda_a(\theta)$ exceeds

$$(1 - \epsilon)\left(m_0 c + \frac{1}{(m_0+1)^2} - 2\epsilon\right) > m_0 c + \frac{1}{(m_0+1)^2} - 3\epsilon. \tag{3.24}$$

Since ϵ may be taken to be arbitrarily small in magnitude, we conclude (see Section 2) that the fixed sample procedure described in Section 2 is indeed minimax.

REFERENCES

[1] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[2] J. Wolfowitz, "Minimax estimates of the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 218–230.

[3] C. Stein and A. Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 427–433.

[4] C. R. Blyth, "On minimax statistical procedures and their admissibility," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 22–42.

EXTENSION OF A METHOD OF INVESTIGATING THE PROPERTIES OF ANALYSIS OF VARIANCE TESTS TO THE CASE OF RANDOM AND MIXED MODELS

By F. N. David and N. L. Johnson

University College, London

Summary. Results are given whereby the methods described in an earlier paper [1], dealing with the parametric case, may be applied also to the case of random, or mixed random and parametric components.

1. Introduction. In a recent paper [1] we set out a method for approximating to the power function of tests of the general linear hypothesis under fairly wide conditions of non-normality and non-uniformity of residual variance. In many analysis of variance problems, it is more reasonable to replace some or all of the parameters by independent random variables with zero expected value. (This is the basis of the well-known 'components of variance' model.)

In the present paper we give certain general formulae which will facilitate the application of the method described in [1] to such random or mixed models. Our results are presented in such a form that they refer to the various sums of squares suggested by the analysis appropriate to the parametric case. Since, however, the same sums of squares are commonly used (though not necessarily in the same way) in the analysis when a random or mixed model is envisaged, the results given will be appropriate in such cases, though care must be taken in their interpretation.

It may be noted that this extension of our method covers the case of the general linear hypothesis with correlated residuals, since such residuals may be represented as the sum of

(i) independent residuals for each observation, and

(ii) independent random terms common to different observations (i.e., occurring in the same way as do parameters in the general linear model).

2. The theoretical model. In [1] we used a theoretical model of the form

$$x_i = a_{i1}\theta_1 + \cdots + a_{i,s-p}\theta_{s-p} + a_{i,s-p+1}\theta_{s-p+1} + \cdots + a_{is}\theta_s + z_i \qquad (i = 1, \cdots, n),$$

where the θ's were unknown parameters and the z's were independent random variables each with zero expected value. The hypothesis to be tested specified that $\theta_{s-p+1} = \cdots = \theta_s = 0$.

We now replace $\theta_{q+1}, \cdots, \theta_{s-p}, \theta_{s-p+r+1}, \cdots, \theta_s (q < s - p, r < p)$ by independent random variables $y_{q+1}, \cdots, y_{s-p}, y_{s-p+r+1}, \cdots, y_s$ (each with expected value zero) so that the theoretical model is of form

$$\begin{aligned} x_i = a_{i1}\theta_1 + \cdots + a_{iq}\theta_q + a_{i,q+1}y_{q+1} + \cdots + a_{i,s-p}y_{s-p} \\ + a_{i,s-p+1}\theta_{s-p+1} + \cdots + a_{i,s-p+r}\theta_{s-p+r} \\ + a_{i,s-p+r+1}y_{s-p+r+1} + \cdots + a_{is}y_s + z_i . \end{aligned}$$

The hypothesis to be tested specifies

$$\theta_{s-p+1} = \cdots = \theta_{s-p+r} = 0,$$
$$\sigma(y_{s-p+r+1}) = \cdots = \sigma(y_s) = 0.$$

As in [1] it is also assumed that the matrix $A = (a_{ij})$ is nonsingular and the z's are mutually independent. We further assume that the y's are independent of the z's.

3. Method of investigation. It will be recalled that in the parametric case the test of the hypothesis $H(\theta_{s-p+1} = \cdots = \theta_s = 0)$ was based on the criterion $(S_b/p)/(S_a/(n-s))$, where S_a is the minimum value of

$$\sum_{i=1}^{n} (x_i - a_{i1}\theta_1 - \cdots - a_{is}\theta_s)^2$$

with respect to $\theta_1, \cdots, \theta_s$; and $S_a + S_b$ is the minimum value of

$$\sum_{i=1}^{n} (x_i - a_{i1}\theta_1 - \cdots - a_{i,s-p}\theta_{s-p})^2$$

with respect to $\theta_1, \cdots, \theta_{s-p}$. The upper $100\alpha\%$ limit of the test criterion could be obtained from tables of significance limits of the F-distribution. The test could formally be expressed as

$$\text{reject } H \text{ if } (S_b/p)/(S_a/(n-s)) > F_{p,n-s,\alpha}.$$

Investigation of properties of the test reduces to evaluation of the probability

$$P\{(S_b/p)/(S_a/(n-s)) > F_{p,n-s,\alpha}\}$$

which can be written in the form

$$P\{S_r - CS_a > 0\},$$

where $C = 1 + pF_{p,n-s,\alpha}/(n-s)$ and $S_r = S_a + S_b$. This probability is obtained approximately by finding a frequency curve which has the same first four moments as $S_r - CS_a$. It is assumed that the theoretical model (1) is adequate in the number of parameters and/or random variables which it contains. Following our previous work it may be shown that S_a and S_r may be written in canonical form as

$$S_a = \sum_{i,j=1}^{n} m_{ij}\, z_i z_j$$

and

$$S_r = \sum_{i,j=1}^{n} m'_{ij}\,(z_i + D'_i)(z_j + D'_j),$$

where

$$D'_i = \sum_{t=s-p+1}^{s-p+r} a_{it}\theta_t + \sum_{t=s-p+r+1}^{s} a_{it} y_t$$
$$= A_i + Y_i \qquad (i = 1, \cdots, n)$$

and the m's and m''s depend only on the a's (see Section 4).

4. Definition of determinants. Before turning to a consideration of the moments it will be convenient to summarise in determinantal form the various quantities which are required.

As before

$$G_{jk} = \sum_{i=1}^{n} a_{ij} a_{ik}$$

and

$$\Delta = \begin{vmatrix} G_{11} & \cdots & G_{1s} \\ \vdots & & \vdots \\ G_{1s} & \cdots & G_{ss} \end{vmatrix}, \qquad \Delta' = \begin{vmatrix} G_{11} & \cdots & G_{1,s-p} \\ \vdots & & \vdots \\ G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix}.$$

Let

$$\Delta'_{ii} = \begin{vmatrix} 1 & a_{i1} & \cdots & a_{i,s-p} \\ a_{i1} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ a_{i,s-p} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix}, \quad \alpha'_{ij} = \begin{vmatrix} 0 & a_{i1} & \cdots & a_{i,s-p} \\ a_{j1} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ a_{j,s-p} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix}.$$

Then

$$m'_{ij} = -\alpha'_{ij}/\Delta' \qquad i \neq j;$$

$$m'_{ii} = 1 - \alpha'_{ii}/\Delta' = \Delta'_{ii}/\Delta'.$$

Similar quantities without primes may be expressed as similar determinants of order $(s + 1)$ instead of $(s - p + 1)$. In this present work we shall also use

$$\delta'_i = \frac{1}{\Delta'} \begin{vmatrix} 0 & \sum_i a_{i1} A_i & \cdots & \sum_i a_{i,s-p} A_i \\ a_{i1} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ a_{i,s-p} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix} = \sum_{j=1}^{n} m'_{ij} A_i,$$

$$\Delta'_A = \frac{1}{\Delta'} \begin{vmatrix} \sum_i A_i^2 & \sum_i a_{i1} A_i & \cdots & \sum_i a_{i,s-p} A_i \\ \sum_i a_{i1} A_i & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ \sum_i a_{i,s-p} A_i & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix} = \sum_{i,j=1}^{n} m'_{ij} A_i A_j.$$

Similar quantities without primes which may be expressed as determinants of order $(s + 1)$ will have zero value. So far the determinants are the same or are directly comparable with those of our previous paper. We now introduce new determinants and note in these definitions that t and u may run from $(s - p + r + 1)$ to s only. We define

$$\Gamma'_{tu} = \frac{1}{\Delta'}\begin{vmatrix} G_{tu} & G_{1t} & \cdots & G_{s-p,t} \\ G_{1u} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ G_{s-p,u} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix}, \quad \Omega'_{it} = \frac{1}{\Delta'}\begin{vmatrix} a_{it} & a_{i1} & \cdots & a_{i,s-p} \\ G_{1t} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ G_{s-p,t} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix},$$

$$\Lambda'_t = \frac{1}{\Delta'}\begin{vmatrix} \sum_i a_{it} A_i & \sum_i a_{i1} A_i & \cdots & \sum_i a_{i,s-p} A_i \\ G_{1t} & G_{11} & \cdots & G_{1,s-p} \\ \vdots & \vdots & & \vdots \\ G_{s-p,t} & G_{1,s-p} & \cdots & G_{s-p,s-p} \end{vmatrix}.$$

Similar determinants of order $(s + 1)$ may be written down to represent quantities without primes but these will be zero.

5. Moments of S_r and S_a. We write

$$\mu(S_r^l S_a^m) = \mathcal{E}[(S_r - \mathcal{E}(S_r))^l (S_a - \mathcal{E}(S_a))^m]$$

with $\kappa(S_r^l S_a^m)$ for the corresponding cumulants. It is easy to see that the moments of S_a are the same as those indicated in [1] with the appropriate determinants now put equal to zero. For example, (all summations running from 1 to n)

$$\mathcal{E}(S_a) = \sum_i m_{ii}\kappa_{2i},$$

$$\kappa(S_a^2) = \sum_i m_{ii}^2 \kappa_{4i} + 2\sum_i \sum_j m_{ij}\kappa_{2i}\kappa_{2j},$$

$$\kappa(S_a^3) = \sum_i m_{ii}^3 \kappa_{6i} + 12\sum_i \sum_j m_{ii} m_{ij}^2 \kappa_{4i}\kappa_{2j} + 6\sum_i \sum_j m_{ii} m_{ij} m_{jj}\kappa_{3i}\kappa_{3j}$$
$$+ 4\sum_i \sum_j m_{ij}^3 \kappa_{3i}\kappa_{3j} + 8\sum_i \sum_j \sum_l m_{ij} m_{il} m_{jl}\kappa_{2i}\kappa_{2j}\kappa_{2l},$$

and so on, the rth cumulant of z_i being defined as κ_{ri} for $r \geqq 2$. Again it is a simple matter to show that $\kappa(S_r S_a^l)$ is the same under this treatment as it was in [1] if the appropriate changes are made in the determinants. Thus

$$\kappa(S_r S_a) = \sum_i m_{ii} m'_{ii}\kappa_{4i} + 2\sum_i \sum_j m_{ij} m'_{ij}\kappa_{2i}\kappa_{2j} + 2\sum_i m_{ii}\delta'_i\kappa_{3i},$$

where δ'_i has A's instead of D's in its definition. The moments of S_r and the cross cumulants of S_a and S_r containing a power of S_r greater than or equal to 2 can be derived by elementary algebra or by a simple combinatorial method from the moments of S_r previously obtained. Let $\bar{\kappa}_{rt}$ be the rth cumulant of y_t.

We have then

$$\mathcal{E}(S_r) = \sum_i m'_{ii}\kappa_{2i} + \Delta'_A + \sum_t \Gamma'_{tt}\bar{\kappa}_{2t},$$

where t may run from $(s - p + r + 1)$ to s only. Again

$$\begin{aligned}\kappa(S_r^2) = \sum_i m'^2_{ii}\kappa_{4i} + 2\sum_i\sum_j m'^2_{ij}\kappa_{2i}\kappa_{2j} + 4\sum_i m'_{ii}\delta'_i\kappa_{2i} + 4\sum_i \delta'^2_i\kappa_{2i} \\ + 4\sum_t \Gamma'^2_{tt}\bar{\kappa}_{4t} + 2\sum_t\sum_u \Gamma'^2_{tu}\bar{\kappa}_{2t}\bar{\kappa}_{2u} + 4\sum_t \Gamma'_{tt}\Lambda'_t\bar{\kappa}_{3t} \\ + 4\sum_t \Lambda'^2_t\bar{\kappa}_{2t} + 4\sum_i\sum_t \Omega'^2_{it}\bar{\kappa}_{2t}\kappa_{2i}.\end{aligned}$$

This last expression demonstrates how the moments of S_r can be obtained directly by substitution from [1]. We write down the expression for $\kappa(S_r^2)$ from [1] and add to it expressions in t, or in t and u, which we obtain by substituting $\bar{\kappa}_{rt}$ for κ_{ri}, Γ'_{tt} for m'_{ii}, and so on. We add further the terms in $\bar{\kappa}_{rt}\kappa_{ri}$ by making the appropriate substitution for the cumulants and writing Ω'_{it} for m'_{ij}. This combinatorial method is obvious if the form of the various determinants is considered. We have worked out the cumulants and cross-cumulants up to and including those of the fourth order by two different methods but they are so easily derived by the above process that we do not reproduce them here in full generality.

6. Special cases of normality. If it is assumed that z_i and y_t are both normally distributed then from a knowledge of the moments it is possible to study the effect of heterogeneity of variance on the power function of the test. For the special case of normality we have

$$\begin{aligned}
\mathcal{E}(S_a) &= \sum m_{ii}\kappa_{2i}, \\
\mathcal{E}(S_r) &= \sum m'_{ii}\kappa_{2i} + \Delta'_A + \sum \Gamma'_{tt}\bar{\kappa}_{2t}, \\
\kappa(S_a^2) &= 2\sum m^2_{ij}\kappa_{2i}\kappa_{2j}, \\
\kappa(S_aS_r) &= 2\sum m_{ij}m'_{ij}\kappa_{2i}\kappa_{2j}, \\
\kappa(S_r^2) &= 2\sum m'^2_{ij}\kappa_{2i}\kappa_{2j} + 4\sum \delta'^2_i\kappa_{2i} + 2\sum \Gamma'^2_{tu}\bar{\kappa}_{2t}\bar{\kappa}_{2u} + 4\sum \Lambda'^2_t\bar{\kappa}_{2t} \\
\kappa(S_a^3) &= 8\sum m_{ij}m_{il}m_{jl}\kappa_{2i}\kappa_{2j}\kappa_{2l}, \\
\kappa(S_rS_a^2) &= 8\sum m_{ij}m_{il}m'_{jl}\kappa_{2i}\kappa_{2j}\kappa_{2l}, \\
\kappa(S_r^2S_a) &= 8\sum m'_{ij}m'_{il}m_{jl}\kappa_{2i}\kappa_{2j}\kappa_{2l} + 8\sum m_{ij}\delta'_i\delta'_j\kappa_{2i}\kappa_{2j} + 8\sum m_{ij}\Omega'_{it}\Omega'_{jt}\kappa_{2i}\kappa_{2j}\bar{\kappa}_{2t}, \\
\kappa(S_r^3) &= 8\sum m'_{ij}m'_{il}m'_{jl}\kappa_{2i}\kappa_{2j}\kappa_{2l} + 8\sum \Gamma'_{tu}\Gamma'_{tv}\Gamma'_{uv}\bar{\kappa}_{2t}\bar{\kappa}_{2u}\bar{\kappa}_{2v} \\
&\quad + 24\sum m'_{ij}\Omega'_{it}\Omega'_{jt}\kappa_{2i}\kappa_{2j}\bar{\kappa}_{2t} + 24\sum \Omega'_{it}\Omega'_{iu}\Gamma'_{tu}\kappa_{2i}\bar{\kappa}_{2t}\bar{\kappa}_{2u} + 24\sum m'_{ij}\delta'_i\delta'_j\kappa_{2i}\kappa_{2j} \\
&\quad + 48\sum \Omega'_{it}\delta'_i\Lambda'_t\kappa_{2i}\bar{\kappa}_{2t} + 24\sum \Gamma'_{tu}\Lambda'_t\Lambda'_u\bar{\kappa}_{2t}\bar{\kappa}_{2u},
\end{aligned}$$

$$
\begin{aligned}
\kappa(S_a^4) &= 48 \sum m_{ij} m_{il} m_{jk} m_{lk} \kappa_{2i} \kappa_{2j} \kappa_{2l} \kappa_{2k} \,, \\
\kappa(S_r S_a^3) &= 48 \sum m'_{ij} m_{il} m_{jk} m_{lk} \kappa_{2i} \kappa_{2j} \kappa_{2l} \kappa_{2k} \,, \\
\kappa(S_r^2 S_a^2) &= 48 \sum m'_{ij} m'_{il} m_{jk} m_{lk} \kappa_{2i} \kappa_{2j} \kappa_{2l} \kappa_{2k} + 48 \sum m_{ij} m_{il} \Omega'_{jt} \Omega'_{lt} \kappa_{2i} \kappa_{2j} \kappa_{2l} \bar{\kappa}_{2t} \\
&\qquad + 32 \sum m_{ij} m_{il} \delta'_j \delta'_l \kappa_{2i} \kappa_{2j} \kappa_{2l} \,, \\
\kappa(S_r^3 S_a) &= 48 \sum m'_{ij} m'_{il} m'_{jk} m_{lk} \kappa_{2i} \kappa_{2j} \kappa_{2l} \kappa_{2k} + 96 \sum m'_{ij} m_{il} \Omega'_{jt} \Omega'_{lt} \kappa_{2i} \kappa_{2j} \kappa_{2l} \bar{\kappa}_{2t} \\
&\qquad + 48 \sum m_{ij} \Omega'_{it} \Omega'_{ju} \Gamma'_{tu} \kappa_{2i} \kappa_{2j} \bar{\kappa}_{2t} \bar{\kappa}_{2u} + 96 \sum m'_{ij} m_{il} \delta'_j \delta'_l \kappa_{2i} \kappa_{2j} \kappa_{2l} \\
&\qquad + 96 \sum m_{ij} \delta'_j \Omega'_{it} \Lambda'_t \kappa_{2i} \kappa_{2j} \bar{\kappa}_{2t} \,, \\
\kappa(S_r^4) &= 48 \sum m'_{ij} m'_{il} m'_{jk} m'_{lk} \kappa_{2i} \kappa_{2j} \kappa_{2l} \kappa_{2k} + 192 \sum m'_{ij} m'_{il} \Omega'_{jt} \Omega'_{lt} \kappa_{2i} \kappa_{2j} \kappa_{2l} \bar{\kappa}_{2t} \\
&\qquad + 288 \sum m'_{ij} \Omega'_{it} \Omega'_{ju} \Gamma'_{tu} \kappa_{2i} \kappa_{2j} \bar{\kappa}_{2t} \bar{\kappa}_{2u} + 192 \sum \Omega'_{it} \Omega'_{iu} \Gamma'_{tv} \Gamma'_{uv} \kappa_{2i} \bar{\kappa}_{2t} \bar{\kappa}_{2u} \bar{\kappa}_{2v} \\
&\qquad + 48 \sum \Gamma'_{tu} \Gamma'_{tv} \Gamma'_{uw} \Gamma'_{vw} \bar{\kappa}_{2t} \bar{\kappa}_{2u} \bar{\kappa}_{2v} \bar{\kappa}_{2w} + 192 \sum m'_{ij} m'_{il} \delta'_j \delta'_l \kappa_{2i} \kappa_{2j} \kappa_{2l} \\
&\qquad + 192 \sum \Omega'_{it} \Omega'_{jt} \delta'_i \delta'_j \kappa_{2i} \kappa_{2j} \bar{\kappa}_{2t} + 384 \sum m'_{ij} \Omega'_{it} \delta'_j \Lambda'_t \kappa_{2i} \kappa_{2j} \bar{\kappa}_{2t} \\
&\qquad + 192 \sum \Omega'_{it} \Omega'_{iu} \Lambda'_t \Lambda'_u \kappa_{2i} \bar{\kappa}_{2t} \bar{\kappa}_{2u} + 384 \sum \Omega'_{it} \Gamma'_{tu} \delta'_i \Lambda'_u \kappa_{2i} \bar{\kappa}_{2t} \bar{\kappa}_{2u} \\
&\qquad + 192 \sum \Gamma'_{tu} \Gamma'_{tv} \Lambda'_u \Lambda'_v \bar{\kappa}_{2t} \bar{\kappa}_{2u} \bar{\kappa}_{2v} \,.
\end{aligned}
$$

For ease of printing each summation sign stands for one, two, three or four separate summations as required by the subscripts. In these summations i, j, l and k run from 1 to n, t, u, v and w run from $(s - p + r + 1)$ to s. A further simplification will be to let δ'_i be zero and the summations for t, u, v and w run from $(s - p)$ to s. In this latter case the alternative hypotheses to that tested specify the existence of certain random variables but not any parameters.

7. Special cases of correlated variables. As an illustration of the use of the foregoing theory when the variables are correlated we consider the test for the linearity of regression in a bivariate table. The standard case where departure from linearity is represented by parameters was studied in [1]. It will now be supposed that the deviations from linearity form a simple moving average series of random variables. Let x_{ti} be the dependent variable and W_t the independent variable ($i = 1, \cdots, n_t$; $t = 1, \cdots, s$). We suppose that the model is

$$x_{ti} = \theta_1 + (W_t - \overline{W})\theta_2 + y_T + R y_{T-1} + z_{ti} \,,$$

where $T = t + 2$ and R is a known constant. We shall assume that $\kappa_r(z_{ti}) = \kappa_{rt}$ (i.e., the distribution of z_{ti} depends only on the array). The fundamental sums of the squares are

$$S_a = \sum_t \sum_i (x_{ti} - \bar{x}_{t.})^2, \qquad S_r = \sum_t \sum_i \{x_{ti} - \bar{x}_{..} - b(W_t - \overline{W})\}^2,$$

where

$$b = \frac{\sum_t n_t (W_t - \overline{W})(\bar{x}_{t.} - \bar{x}_{..})}{\sum_t n_t (W_t - \overline{W})^2} \,.$$

Evaluation of the determinants gives

$$\Gamma'_{TT} = n_t + R^2 n_{t+1} - N^{-1}(n_t + Rn_{t+1})^2 - (\sum n_t w_t^2)^{-1}(n_t w_t + Rn_{t+1}w_{t+1})^2,$$

where $w_t = W_t - \overline{W}$ and $n_{s+1} = 0$. We have, therefore, using the determinants α_{ij} which have been worked out in [1],

$$\mathcal{E}(S_r) = \sum n_t\left(1 - \frac{1}{N} - \frac{w_t^2}{\sum n_t w_t^2}\right)\kappa_{2t}$$

$$+ \sum\left[n_t + R^2 n_{t+1} - \frac{(n_t + Rn_{t+1})^2}{N} - \frac{(n_t w_t + Rn_{t+1} w_{t+1})^2}{\sum n_t w_t^2}\right]\bar{\kappa}_{2T}$$

with the convention that $n_{s+1} = 0$. Again it may be shown that

$$\Gamma'_{T,T+1} = Rn_{t+1} - \frac{(n_t + Rn_{t+1})(n_{t+1} + Rn_{t+2})}{N} - \frac{(n_t w_t + Rn_{t+1} w_{t+1})(n_{t+1} w_{t+1} + Rn_{t+2} w_{t+2})}{\sum n_t w_t^2}$$

and

$$\Gamma'_{TU} = -\frac{(n_t + Rn_{t+1})(n_u + Rn_{u+1})}{N} - \frac{(n_t w_t + Rn_{t+1} w_{t+1})(n_u w_u + Rn_{u+1} w_{u+1})}{\sum n_t w_t^2}$$

where $U = u + 2$ $(u = 1, \cdots s)$ and $| T - U | = | t - u | > 1$. Also if in terms of our original notation (Sections 2–6) i is in the tth group,

$$\Omega'_{iT} = 1 - \frac{n_t + Rn_{t+1}}{N} - \frac{w_t(n_t w_t + Rn_{t+1} w_{t+1})}{\sum n_t w_t^2},$$

if i is in the $(t + 1)$th group,

$$\Omega'_{iT} = R - \frac{n_t + Rn_{t+1}}{N} - \frac{w_t(n_t w_t + Rn_{t+1} w_{t+1})}{\sum n_t w_t^2},$$

and if i is not in the tth or the $(t + 1)$th groups,

$$\Omega'_{iT} = -\frac{n_t + Rn_{t+1}}{N} - \frac{w_t(n_t w_t + Rn_{t+1} w_{t+1})}{\sum n_t w_t^2}.$$

For brevity we write

$$\phi_{tu} = \frac{1}{N} + \frac{w_t w_u}{\sum n_t w_t^2},$$

$$\chi_{tu} = \frac{(n_t + Rn_{t+1})(n_u + Rn_{u+1})}{N} + \frac{(n_t w_t + Rn_{t+1} w_{t+1})(n_u w_u + Rn_{u+1} w_{u+1})}{\sum n_t w_t^2},$$

$$\psi_t = \frac{n_t + Rn_{t+1}}{N} + \frac{w_t(n_t w_t + Rn_{t+1} w_{t+1})}{\sum n_t w_t^2}.$$

Then

$$\begin{aligned}\kappa(S_r^2) = &\sum n_t(1 - \phi_{tt})^2\kappa_{4t} + \sum n_t(1 - 2\phi_{tt})\kappa_{2t}^2 + 2\sum n_t n_u \phi_{tu}\kappa_{2t}\kappa_{2u} \\ &+ \sum (n_t + R^2 n_{t+1} - \chi_{tt})^2 \bar{\kappa}_{4T} + 2\sum (n_t + R^2 n_{t+1})(n_t + R^2 n_{t+1} - 2\chi_{tt})\bar{\kappa}_{2T}^2 \\ &+ 4\sum R n_{t+1}(R n_{t+1} - 2\chi_{t,t+1})\bar{\kappa}_{2T}\bar{\kappa}_{2,T+1} + 2\sum \chi_{tu}^2\kappa_{2T}\kappa_{2U} \\ &+ \sum [n_t(1 - 2\psi_t)\kappa_{2t} + n_{t+1}R(1 - 2R\psi_t)\kappa_{2,t+1} + \sum n_u\psi_u\kappa_{2u}]\bar{\kappa}_{2T}.\end{aligned}$$

The higher cumulants follow in a similar way.

REFERENCES

[1] F. N. DAVID AND N. L. JOHNSON, "A method of investigating the effect of nonnormality and heterogeneity of variance on tests of the general linear hypothesis," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 382–392.

SOME RELATIONS AMONG THE BLOCKS OF SYMMETRICAL GROUP DIVISIBLE DESIGNS

By W. S. Connor

National Bureau of Standards[1]

1. Summary. It is well known that if every pair of treatments in a symmetrical balanced incomplete block design occurs in λ blocks, then every two blocks of the design have λ treatments in common. In this paper it will be shown that a somewhat similar property holds for symmetrical group divisible designs. In the course of the investigation there will be introduced certain matrices which are of intrinsic interest.

2. Introduction. Some of the combinatorial properties of group divisible incomplete block designs were considered in [1]. Here we shall need the definition of group divisible designs and the three classes into which they fall. An incomplete block design with v treatments each replicated r times in b blocks of size k is said to be group divisible (GD) if the treatments can be divided into m groups, each with n treatments, so that the treatments belonging to the same group occur together in λ_1 blocks and the treatments belonging to different groups occur together in λ_2 blocks, $\lambda_1 \neq \lambda_2$. The three exhaustive and mutually exclusive classes into which the GD designs fall are as follows:

(a) Singular GD designs characterized by $r - \lambda_1 = 0$;
(b) Semi-regular GD designs characterized by $r - \lambda_1 > 0$, $rk - v\lambda_2 = 0$; and
(c) Regular GD designs characterized by $r - \lambda_1 > 0$, $rk - v\lambda_2 > 0$.

In this paper we shall study classes (b) and (c) for the symmetrical case, that is, the case when $r = k$, or equivalently, $b = v$.

3. The incidence and structural matrices. In [2] there was defined the structural matrix for balanced incomplete block designs. We now shall define the incidence matrix, and two structural matrices for *GD* designs.

Let us consider first the incidence matrix of a *GD* design,

$$N = \begin{bmatrix} n_{11} & \cdots\cdots & n_{1b} \\ \vdots & & \vdots \\ n_{v1} & \cdots\cdots & n_{vb} \end{bmatrix}, \tag{3.1}$$

where the rows represent treatments, the columns represent blocks, and $n_{ij} = 1$ or 0 according as the ith treatment does or does not occur in the jth block. From the conditions satisfied by the design it is easy to see that

$$\sum_{j=1}^{b} n_{ij} = r \qquad (i = 1, \cdots, v), \tag{3.2}$$

[1] This work was begun while the author was at the University of North Carolina.

and

$$\sum_{j=1}^{b} n_{ij} n_{uj} = \lambda_1 \quad \text{or} \quad \lambda_2, \tag{3.3}$$

according as the ith and uth treatments ($i \neq u$) do belong or do not belong to the same group.

Throughout the paper let us adopt the convention that the treatments $n(w - 1) + 1, n(w - 1) + 2, \cdots, nw$ shall belong to the wth group ($w = 1, \cdots, m$). Then

$$NN' = \begin{bmatrix} A & B \cdots B \\ B & A \cdots B \\ \cdot & \cdot \quad\; \cdot \\ \cdot & \cdot \quad\; \cdot \\ \cdot & \cdot \quad\; \cdot \\ B & B \cdots A \end{bmatrix}, \tag{3.4}$$

where the elements of the nxn submatrix A are r in the principal diagonal and λ_1 elsewhere, and the elements of the nxn submatrix B are λ_2 everywhere. Of course NN' contains $v = mn$ rows and columns.

Now choose any $t \leqq b$ blocks of the design. Let the submatrix of N which corresponds to these t blocks be denoted by N_0. Let s_{ju} be the number of treatments common to the jth and uth chosen blocks ($j, u = 1, 2, \cdots, t$). Then the $t \times t$ symmetric matrix

$$S_t^I = N_0' N_0 = (s_{ju}) \tag{3.5}$$

is defined to be the *intersection structural matrix of the t chosen blocks*. The jth row or column of S_t^I corresponds to the jth chosen block and the successive elements of the jth row or column give the number of treatments which this block has in common with the 1st, 2nd, $\cdots$, tth chosen blocks.

We next shall consider another structural matrix. Let $\overset{w}{s}_{ju}$ denote the number of treatments from the wth group which blocks j and u have in common. Then

$$\sum_{w=1}^{m} \overset{w}{s}_{ju} = s_{ju}, \tag{3.6}$$

$$\sum_{w=1}^{m} \overset{w}{s}_{jj} = k. \tag{3.7}$$

Now consider the matrix

$$G_t = \begin{bmatrix} \overset{1}{s}_{11} & \overset{1}{s}_{22} & \cdots & \overset{1}{s}_{tt} \\ \overset{2}{s}_{11} & \overset{2}{s}_{22} & \cdots & \overset{2}{s}_{tt} \\ \cdot & \cdot & \cdots & \cdot \\ \overset{m}{s}_{11} & \overset{m}{s}_{22} & \cdots & \overset{m}{s}_{tt} \end{bmatrix} \tag{3.8}$$

and the product matrix

$$S_t^G = G_t' G_t, \tag{3.9}$$

where the element in the jth row and the uth column is the sum of products of the number of treatments which the jth chosen block and the uth chosen block contain from each group. We define S_t^G as the *group structural matrix of the t chosen blocks.*

4. The characteristic matrix. We shall define an analogue of the characteristic matrix which was developed for balanced incomplete block designs in [2]. For the remainder of the paper, except for the last section, we shall restrict our attention to the regular GD designs.

Let the columns of N be permuted so that the first t columns correspond to the t chosen blocks. Then let the incidence matrix be extended by adjoining t new rows, so that the elements of the jth adjoined row are zero, except for the jth which is unity. We thus get

$$N_1 = \begin{bmatrix} N & \\ I_t & 0 \end{bmatrix}, \tag{4.1}$$

where I_t is the identity matrix of order t, and 0 is the $t \times (b - t)$ zero matrix. Then

$$N_1 N_1' = \begin{bmatrix} NN' & N_0 \\ N_0' & I_t \end{bmatrix}. \tag{4.2}$$

The evaluation of $| N_1 N_1' |$ leads to

$$| N_1 N_1' | = (rk)^{-t+1}(r - \lambda_1)^{v-t-m}(rk - v\lambda_2)^{m-t-1} \, | C_t | , \tag{4.3}$$

where the typical element of C_t is

$$c_{ju} = (rk - v\lambda_2)(rk\delta_{ju} + \lambda_2 k^2) + (\lambda_1 - \lambda_2)\left(rk \sum_{w=1}^{m} s_{jj}^{w} s_{uu}^{w} - n\lambda_2 k^2\right), \tag{4.4}$$

where $\delta_{ju} = (r - \lambda_1 - k)$ or $-s_{ju}$, according as $j = u$ or $j \neq u$. The matrix C_t is defined as the *characteristic matrix of the t chosen blocks.* The jth row or the jth column of C_t corresponds to the jth chosen block of the design.

We observe that the characteristic matrix is related to the two structural matrices as is described in the following theorem.

THEOREM 4.1. *For the regular GD designs there exists a (1–1) correspondence among the elements of the intersection structural matrix S_t^I, the group structural matrix S_t^G, and the characteristic matrix C_t. This correspondence is given by*

$$C_t = rk(rk - v\lambda_2)[(r - \lambda_1)I_t - S_t^I] + rk(\lambda_1 - \lambda_2)S_t^G + \lambda_2 k^2 (r - \lambda_1)E_t ,$$

where E_t is the singular $t \times t$ matrix all of whose elements are unity.

For the particular case when $r = k$, the value of $| N_1 N_1' |$ as given by (4.3) reduces to

$$| N_1 N_1' | = r^{-2(t-1)}(r - \lambda_1)^{v-t-m}(r^2 - v\lambda_2)^{m-t-1} \, | C_t | , \tag{4.5}$$

where the typical element of C_t is

$$c_{ju} = r^2(r^2 - v\lambda_2)(\delta_{ju} + \lambda_2) + r^2(\lambda_1 - \lambda_2)\left(\sum_{w=1}^{m} s_{jj}^{w} s_{uu}^{w} - n\lambda_2\right). \tag{4.6}$$

We shall state an analogue of Theorem 3.1 of [2]. The proof is as for that theorem.

THEOREM 4.2. *If C_t is the characteristic matrix of any set of t blocks chosen from a regular GD design with parameters v, b, r, k, m, n, λ_1, and λ_2, then*

(i) $| C_t | \geqq 0$ *if* $t < b - v$,

(ii) $| C_t | = 0$ *if* $t > b - v$, *and*

(iii) $r^{-2(t-1)}(r - \lambda_1)^{v-t-m}(r^2 - v\lambda_2)^{m-t-1} \mid C_t \mid$ is a perfect integral square, if $t = b - v$.

5. Inequalities on s_{ju} for regular symmetrical designs. Let $t = 1$. Then since the factor outside of $| C_1 |$ in (4.5) is positive, it follows from Theorem 4.2 that $| C_1 | = 0$. Hence, from (4.6),

$$r^2(\lambda_1 - \lambda_2)\left[\sum_{w=1}^{m} (s_{11}^{w})^2 - r^2 + v\lambda_2 - n\lambda_2\right] = 0. \tag{5.1}$$

Since $r^2(\lambda_1 - \lambda_2) \neq 0$,

$$\sum_{w=1}^{m} (s_{11}^{w})^2 = r^2 - v\lambda_2 + n\lambda_2. \tag{5.2}$$

Now let $t = 2$. Since $c_{11} = c_{22} = 0$, it is necessary by Theorem 4.2 that $c_{12} = c_{21} = 0$. Hence from (4.6),

$$s_{12} = \lambda_2 + \frac{e}{(r^2 - v\lambda_2)}(\lambda_1 - \lambda_2), \tag{5.3}$$

where

$$e = \sum_{w=1}^{m} s_{11}^{w} s_{22}^{w} - n\lambda_2.$$

From (5.2) and the observation that $s_{jj}^{w} \geqq 0$ ($j = 1, 2; w = 1, \cdots, m$), it follows that

$$-n\lambda_2 \leqq e \leqq r^2 - v\lambda_2. \tag{5.4}$$

From (5.3) and (5.4) we obtain

THEOREM 5.1. *For a regular symmetrical GD design the number of treatments s_{ju} common to two blocks satisfies the inequalities*

$$\lambda_2(r - \lambda_1)/(r^2 - v\lambda_2) \leqq s_{ju} \leqq \lambda_1,$$

when $\lambda_1 > \lambda_2$. The inequalities are reversed when $\lambda_1 < \lambda_2$.

6. The block structure for regular symmetrical GD designs when $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ are relatively prime. We need to consider the distribution of the treatments contained in an initial block B_1 among the other blocks. Let n_j be the number of blocks among the remaining $(b - 1)$ blocks which have j treatments in common with B_1. Then from the definition of the design we obtain

$$\begin{aligned} \sum_{j=0}^{k} n_j &= b - 1 = v - 1, \\ \sum_{j=0}^{k} jn_j &= r(k - 1) = r(r - 1). \end{aligned} \tag{6.1}$$

Also consider $M = \sum_{j=0}^{k} j(j - 1)n_j$, which is twice the number of pairs of treatments of B_1 which lie among the other blocks. M is given by

$$M = \sum_{w=1}^{m} s_{11}^{w}(s_{11}^{w} - 1)(\lambda_1 - 1) + \sum_{\substack{x,w=1 \\ x\neq w}}^{m} s_{11}^{x} s_{11}^{w}(\lambda_2 - 1). \tag{6.2}$$

From (3.7) and (5.2), since $r = k$,

$$\sum_{w=1}^{m} s_{11}^{w}(s_{11}^{w} - 1) = (n - 1)\lambda_1, \tag{6.3}$$

$$\sum_{\substack{x,w=1 \\ x\neq w}}^{m} s_{11}^{x} s_{11}^{w} = (m - 1)n\lambda_2. \tag{6.4}$$

Hence

$$M = (n - 1)(\lambda_1)(\lambda_1 - 1) + (m - 1)(n)(\lambda_2)(\lambda_2 - 1). \tag{6.5}$$

Now consider

$$B = \sum_{j=0}^{k} (j - \lambda_1)(j - \lambda_2)n_j. \tag{6.6}$$

From (6.1), (6.5), and (6.6) we obtain

$$B = 0. \tag{6.7}$$

Hence the following lemma.

LEMMA 6.1. *If for a regular symmetrical GD design n_j denotes the number of blocks which have j treatments in common with a given initial block, then*

$$B = \sum_{j=0}^{k} n_j(j - \lambda_1)(j - \lambda_2) = 0.$$

Now let $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ be relatively prime. It follows from (5.3) that s_{12} cannot lie in the open interval (λ_1, λ_2). Then every term of B is positive or zero. But since $B = 0$, every term must be zero. We thus get

THEOREM 6.1. *If for a regular symmetrical GD design $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ are relatively prime, then any two blocks have either λ_1 or λ_2 treatments in common.*

We further observe that even if $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ are not relatively prime, it still may not be possible to choose the elements of G_t of (3.8), subject to the restrictions of (3.7) and (5.2), such that s_{ju} is integral, but is not λ_1 or λ_2. Consider, for example, the *GD* design with parameters $v = b = 45$, $r = k = 9$, $m = 3$, $n = 15$, $\lambda_1 = 3$, and $\lambda_2 = 1$. The highest common factor of $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ is 2. It is clear that the only positive integers which satisfy (3.7) and (5.2) are 1, 1, and 7. But then we must have either $\sum_{w=1}^{m} s_{jj}^{w} s_{uu}^{w} = 51$ or 15, which correspond respectively to λ_1 and λ_2.

Now assume that the condition of Theorem 6.1 is met, or more generally, that positive integers do not exist which meet the restrictions of (3.7), (5.2) and Lemma 6.1 and imply values of s_{ju} other than λ_1 and λ_2. Then from (6.1) we obtain

$$\begin{aligned} n_{\lambda_1} + n_{\lambda_2} &= v - 1, \\ \lambda_1 n_{\lambda_1} + \lambda_2 n_{\lambda_2} &= r(r - 1), \end{aligned} \tag{6.8}$$

whence

$$\begin{aligned} n_{\lambda_1} &= n - 1, \\ n_{\lambda_2} &= (m - 1)n, \end{aligned} \tag{6.9}$$

so that with respect to any initial block B_1, there are $(n - 1)$ other blocks which have λ_1 treatments in common with it, and $(m - 1)n$ other blocks which have λ_2 treatments in common with it.

From (5.3) we see that

$$\sum_{w=1}^{m} s_{11}^{w} s_{jj}^{w} = r + (n - 1)\lambda_1 \tag{6.10}$$

implies that blocks 1 and j have λ_1 treatments in common, and conversely. But then from (5.2) and (6.10), it follows that

$$\sum_{w=1}^{m} s_{11}^{w} s_{jj}^{w} = \sum_{w=1}^{m} (s_{11}^{w})^2, \tag{6.11}$$

which implies that $s_{11}^{w} = s_{jj}^{w}$, $(w = 1, \cdots, m; j = 2, \cdots, b)$. Hence, if blocks B_1 and B_j have λ_1 treatments in common, and blocks B_1 and B_u have λ_1 treatments in common, then B_j and B_u have λ_1 treatments in common. We thus have

THEOREM 6.2. *If for a regular symmetrical GD design $r^2 - v\lambda_2$ and $\lambda_1 - \lambda_2$ are relatively prime, then the blocks fall into m groups of n blocks each, which are such that any two blocks from the same group contain λ_1 treatments in common and any two blocks from different groups contain λ_2 treatments in common.*

As has been indicated above, this theorem could be stated somewhat more generally.

7. The semi-regular class. For this class $rk - v\lambda_2 = 0$, and hence the above theory does not apply. We shall give a simple example which demonstrates for small v that there do sometimes exist solutions in which $s_{ju} \neq \lambda_1$ or λ_2 for some j and u.

Consider the GD design with parameters $v = b = 8$, $r = k = 4$, $m = 4$, $n = 2$, $\lambda_1 = 0$, and $\lambda_2 = 2$. One solution is

$$N^{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

which has the property that the blocks break up into 4 groups of 2 blocks each, which are such that two blocks in the same group have zero treatments in common and any two blocks from different groups have 2 treatments in common.

Another solution is

$$N^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

which is such that any initial block has 1 treatment in common with each of three blocks, 2 treatments in common with each of three blocks, and 3 treatments in common with one block.

We shall now obtain inequalities for the number of treatments s_{ju} in common to any two blocks of a symmetrical semiregular GD design. Since for a semiregular GD design, $rk = v\lambda_2$, it follows that $r - \lambda_1 = n(\lambda_2 - \lambda_1)$, from which we obtain the following lemma.

LEMMA 7.1. *For a semi-regular GD design, it is necessary that* $\lambda_2 > \lambda_1$.

Now let $r = k$. Choose any two blocks and let the columns of N be permuted so that the first two columns correspond to the chosen blocks. Then to N affix m new columns, the wth of which contains $(\lambda_2 - \lambda_1)^{\frac{1}{2}}$ in the rows which correspond to the treatments of the wth group, $(w = 1, \cdots, m)$, and zero elsewhere. Let the augmented matrix be denoted by N_2. Now form

$$N_3 = \begin{bmatrix} & N_2 & \\ I_2 & & 0 \end{bmatrix}, \tag{7.1}$$

where I_2 is the identity matrix of order 2 and 0 is the $2x(b + m - 2)$ matrix all of whose elements are zero. Then

$$|N_3N_3'| = (r + v\lambda_2 - \lambda_1)^{-1}(r - \lambda_1)^{v-3}|B_2|, \tag{7.2}$$

where B_2 is a 2×2 matrix with elements

$$\begin{aligned} b_{11} = b_{22} &= (r + v\lambda_2 - \lambda_1)(-\lambda_1) + \lambda_2 r^2, \\ b_{12} = b_{21} &= (r + v\lambda_2 - \lambda_1)(-s_{12}) + \lambda_2 r^2. \end{aligned} \tag{7.3}$$

As for Theorem 4.2 it is necessary that $|N_3N_3'| \geqq 0$, and since the factor outside of $|B_2|$ in (7.2) is positive, it is necessary that $|B_2| \geqq 0$. Hence, the following theorem:

THEOREM 7.1. *For a symmetrical semi-regular GD design, the number of treatments common to two blocks, s_{ju}, satisfies the inequalities*

$$\lambda_1 \leqq s_{ju} \leqq \frac{2\lambda_2 r^2}{r + v\lambda_2 - \lambda_1} - \lambda_1.$$

I wish to express my thanks to Professor R. C. Bose for suggesting this problem.

REFERENCES

[1] R. C. BOSE AND W. S. CONNOR, "Combinatorial properties of group divisible incomplete block designs." *Annals of Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[2] W. S. CONNOR, JR., "On the structure of balanced incomplete block design." *Annals of Math. Stat.*, Vol. 23 (1952), pp. 57–71.

AN OPTIMUM SOLUTION TO THE k-SAMPLE SLIPPAGE PROBLEM FOR THE NORMAL DISTRIBUTION[1]

By Edward Paulson

University of Washington

0. Summary. A slippage problem for normal distributions is formulated as a multiple decision problem, and a solution is obtained which has certain optimum properties. The discussion is confined to the fixed sample case with the same number of observations from each distribution, and the normal distributions involved are assumed to have a common but unknown variance.

1. Introduction. This paper will consider the problem of how to compare k categories, such as k varieties of wheat, k machines, k teaching methods, etc., so as to decide on the basis of a random sample of n observations with each category whether or not the categories are equal, and if not which is the 'best' one. A problem of this type has been discussed by Mosteller [1] for the nonparametric case. In previous papers [2], [3], we had considered some different types of multiple-decision problems arising in the comparison of k categories, and the emphasis had been on studying the distribution problems involved when the statistical procedures used were suggested by intuitive considerations. In this paper we will be primarily concerned with finding a statistical procedure which in some reasonable sense is an 'optimum' one.

In this paper we will restrict our attention to the case where the n observations $x_{i1}, x_{i2}, \cdots, x_{in}$ in the ith category Π_i are assumed to be normally and independently distributed with mean m_i and a common standard deviation σ, and the best category is (for convenience) defined to be the one associated with the greatest mean value. Let D_0 denote the decision that the k means are all equal, and let D_j $(j = 1, 2, \cdots, k)$ denote the decision that D_0 is incorrect and $m_j = \max(m_1, m_2, \cdots, m_k)$. Our problem is to find a statistical procedure for choosing one of the $k + 1$ decisions $(D_0, D_1, \cdots, D_k)$ which will be in some sense an optimum one. At this point, instead of introducing a weight function as required in the general theory as developed by Wald [4], we will follow a simpler plan which is somewhat analogous to the classical Neyman-Pearson theory of testing a hypothesis, and attempt to find a statistical procedure which, subject to certain restrictions, will in certain instances maximize the probability of making the correct decision.

In order to give a more precise formulation and the solution to the problem let $x_{i\alpha}$ denote the αth observation in the sample from Π_i $(i = 1, 2, \cdots, k;$ $\alpha = 1, 2, \cdots, n)$, let $\bar{x}_i = \sum_{\alpha=1}^{n} (x_{i\alpha}/n)$, $\bar{x} = \sum_{i=1}^{k} (\bar{x}_i/k)$, $s^2 = \sum_{i=1}^{k} \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)^2/[k(n-1)]$, and let M be the subscript of the category with the greatest sample mean, so that $\bar{x}_M = \max\{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$. We will say that the category Π_i has slipped to the right by an amount Δ $(\Delta > 0)$ if $m_1 = m_2 =$

[1] Work done under the sponsorship of the Office of Naval Research.

$\cdots = m_{i-1} = m_{i+1} = \cdots = m_k$ and $m_i = m_1 + \Delta$. The first formulation of the problem is the following: to find a statistical procedure for selecting one of the decisions $(D_0, D_1, \cdots, D_k)$ which will maximize the probability of making the correct decision when some category has slipped to the right subject to the restriction (a) when all the means are equal, D_0 should be selected with probability $1 - \alpha$ (where α is some small positive number fixed in advance of the experiment). In this formulation, the class of allowable decision procedures seems to be too large to admit of an optimum solution and we will, therefore, limit the class of allowable statistical procedures by the following additional restrictions. (b) The decision procedure must be invariant if a constant is added to all the observations, (c) the decision procedure must be invariant when all the observations are multiplied by a positive constant, and (d) the decision procedure must be symmetric in the sense that the probability of making the correct decision when category Π_i has slipped to the right by an amount Δ must be the same for $i = 1, 2, \cdots, k$. These additional restrictions are rather weak and seem to be reasonable requirements to impose in many practical problems. The problem is now reformulated as follows: to find a statistical procedure for selecting one of the set $(D_0, D_1, \cdots, D_k)$ which, subject to restrictions (a), (b), (c) and (d), will maximize the probability of making the correct decision when one of the categories has slipped to the right. The optimum solution will be shown to be the following procedure:

$$(1)\qquad \begin{aligned} &\text{if } \frac{n(\bar{x}_M - \bar{x})}{\sqrt{\sum_{i=1}^{k}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x})^2}} > \lambda_\alpha, \text{ select } D_M;\\ &\text{if } \frac{n(\bar{x}_M - \bar{x})}{\sqrt{\sum_{i=1}^{k}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x})^2}} \leqq \lambda_\alpha, \text{ select } D_0. \end{aligned}$$

Here λ_α is a constant whose precise value is determined by requirement (a), and does not depend on Δ or σ. Since for a given k and n the value of λ_α depends only on α, the optimum property of (1) holds uniformly in Δ and σ.

2. Derivation of the optimum procedure. There is obviously no loss of generality in only considering statistical procedures which depend on the set $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k, s^2)$ since these constitute a set of sufficient statistics for the unknown parameters $(m_1, m_2, \cdots, m_k, \sigma^2)$. Making use of this in connection with restrictions (b) and (c) it is easy to see that any allowable decision procedure will depend only on the $k-1$ statistics $(\bar{x}_1 - \bar{x}_k)/s, (\bar{x}_2 - \bar{x}_k)/s, \cdots, (\bar{x}_{k-1} - \bar{x}_k)/s$. Let $w_\alpha = (\bar{x}_\alpha - \bar{x}_k)/s$ and let $a_\alpha = (m_\alpha - m_k)/\sigma$ for $\alpha = 1, 2, \cdots, k-1$. The joint probability distribution of the set $(w_1, w_2, \cdots, w_{k-1})$ depends only on the parameters $(a_1, a_2, \cdots, a_{k-1})$. Let $\bar{D}_0$ denote the decision tha $a_1 = a_2 = \cdots = a_{k-1} = 0$, and for $1 \leqq j \leqq k - 1$ let $\bar{D}_j$ denote the decision that $a_1 = a_2 = \cdots = a_{j-1} = a_{j+1} = \cdots = a_{k-1} = 0$ and $a_j = \Delta/\sigma$, while $\bar{D}_k$ denotes the decision that $a_1 = a_2 = \cdots = a_{k-1} = -\Delta/\sigma$. Since any allowable decision procedure for selecting one of the set $(D_0, D_1, \cdots, D_k)$ must be a function only of $(w_1, w_2, \cdots, w_{k-1})$ it can be transformed in a natural manner into a decision procedure for selecting one of the decisions $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ by making D_i correspond to $\bar{D}_i$ for $i = 0, 1, 2, \cdots, k$; that is, whenever the original decision procedure selects D_i the transformed decision procedure is to select $\bar{D}_i$. Because of restriction (a), the probability that any transformed allowable decision procedure will select $\bar{D}_0$ when $a_1 = a_2 = \cdots = a_{k-1} = 0$ will be equal to $1 - \alpha$; in addition the probability that any allowable decision procedure will select D_i when Π_i has slipped to the right by an amount Δ is equal to the probability that the transformed procedure select $\bar{D}_i$ when $\bar{D}_i$ is the correct decision, and this last probability must be the same for each i because of restriction (d).

The proof that (1) is the optimum solution consists mainly in showing that for any Δ and σ there exist a set of nonzero a priori probabilities $g_0, g_1, \cdots, g_k$ which are functions of Δ and σ so that when (1) is transformed in the manner indicated above into a decision procedure for selecting one of $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$, it will maximize the probability of making the correct decision among the set $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ when g_i is the a priori probability that $\bar{D}_i$ is the correct decision. Assuming this has been demonstrated, it follows easily that (1) must be the optimum solution. For suppose there existed an allowable decision procedure D^*, which for some Δ and σ had a greater probability than (1) of making the correct decision when some category had slipped to the right by an amount Δ. Then D^*, which must be a function only of $(w_1, w_2, \cdots, w_{k-1})$ when transformed in the indicated manner into a decision procedure for selecting one of $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ will have a greater probability than (1) of making the correct decision among $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ with respect to any set of non-zero a priori probabilities, which would be a contradiction.

To show that the required a priori distribution exists, first let $u_\alpha = (\bar{x}_\alpha - \bar{x}_k)/\sigma$ $(\alpha = 1, 2, \cdots, k - 1)$ so that $w_\alpha = (u_\alpha \sigma / s)$. The random variables $(u_1, u_2, \cdots, u_{k-1})$ can easily be verified to have a $(k - 1)$ dimensional multivariate normal distribution with common variance $= 2/n$, common correlation $= \frac{1}{2}$, and mean values $(a_1, a_2, \cdots, a_{k-1})$. By an elementary calculation, the joint probability density function of $u_1, u_2, \cdots, u_{k-1}$ is given by $C_1 \exp [-\frac{1}{2}\{A \sum_{\alpha=1}^{k-1} (u_\alpha - a_\alpha)^2 + B \sum_{\alpha \neq \beta} (u_\alpha - a_\alpha)(u_\beta - a_\beta)\}]$ where $A = ((k - 1)n/k)$, $B = -n/k$, and C_1 is a constant whose precise value is not needed. Using this result plus the known facts that $n's^2/\sigma^2$ has the χ^2 distribution with $n' = k(n - 1)$ degrees of freedom and is independent of the set $u_1, u_2, \cdots, u_{k-1}$, the joint probability density function $f(w_1, w_2, \cdots, w_{k-1})$ of $w_1, \cdots, w_{k-1}$ is easily found to be given by

$$f(w_1, w_2, \cdots, w_{k-1}) = C_2 \int_0^\infty y^{n'+k-2} \exp \left[\quad \left\{ n' y^2 + A \sum_{\alpha=1}^{k-1} (w_\alpha y - a_\alpha)^2 + B \sum_{\alpha \neq \beta} (w_\alpha y - a_\alpha)(w_\beta y - a_\beta) \right\} \right] dy. \tag{2}$$

Let $f_i = f(w_1, \cdots, w_{k-1} \mid \bar{D}_i)$ denote the joint probability density function of $w_1, \cdots, w_{k-1}$ when $\bar{D}_i$ is the correct decision. The decision procedure which will maximize the probability of making the correct decision among the set $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ when the a priori probability distribution is $(p_0, p_1, p_2, \cdots, p_k)$, that is, the Bayes solution with respect to $(p_0, p_1, \cdots, p_k)$, is known [4] to be given by the rule: for each j $(j = 0, 1, \cdots, k)$ select $\bar{D}_j$ for all points in the $w_1 \cdots w_{k-1}$ space where $p_j f_j = \max\{p_0 f_0, p_1 f_1, \cdots, p_k f_k\}$. For the problem at hand, this is the unique Bayes solution except possibly for a set of measure zero according to all f_i. Using (2) it is easy to calculate for each j the region where $\bar{D}_j$ is selected for the special a priori distribution $p_0 = (1 - kp)$, $p_1 = p_2 \cdots = p_k = p$. For example the region where $\bar{D}_1$ is selected is given by the points in the w space where $f_1 > f_2, f_1 > f_3, \cdots, f_1 > f_k$, and $pf_1 > (1 - kp)f_0$. For any j with $1 < j < k$, the region where $f_1 > f_j$ is given by

$$\int_0 y^{n'+k-2} \exp\left[-\tfrac{1}{2}\left(n' y^2 + Ay^2 \sum_{\alpha=1}^{k-1} w_\alpha^2 + By^2 \sum_{\alpha \neq \beta} w_\alpha w_\beta + A \frac{\Delta^2}{\sigma^2} - 2B \frac{\Delta}{\sigma} y \sum_{\alpha=1}^{k-1} w_\alpha\right)\right]$$
$$\cdot \left\{\exp\left[A \frac{\Delta}{\sigma} yw_1 - B \frac{\Delta}{\sigma} yw_1\right] - \exp\left[A \frac{\Delta}{\sigma} yw_j - B \frac{\Delta}{\sigma} yw_j\right]\right\} dy > 0.$$

The integrand is positive for all y in the range $0 < y < \infty$ if $w_1 > w_j$, and the integrand is negative for all y in this range when $w_1 < w_j$, (since $A - B > 0$) so that $f_1 > f_j$ for $1 < j < k$ if and only if $w_1 > w_j$. In a similar manner, it is easy to show that $f_1 > f_k$ if and only if $w_1 > 0$. The region where $pf_1 > (1 - kp)f_0$ is given by

$$\int_0^\infty y^{n'+k-2} \exp\left[-\frac{y^2}{2}\left(n' + A \sum_{\alpha=1}^{k-1} w_\alpha^2 + B \sum_{\alpha \neq \beta} w_\alpha w_\beta\right)\right]$$
$$\cdot \left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left[\left\{(A - B)\frac{\Delta}{\sigma} w_1 + B \frac{\Delta}{\sigma} \sum_{\alpha=1}^{k-1} w_\alpha\right\} y\right] - (1 - kp)\right\} dy > 0.$$

Making a change of variable, this region is equivalent to

$$\int_0^\infty t^{n'+k-2} \exp\left(-\frac{t^2}{2}\right)\left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left[\frac{\Delta}{\sigma} h(w_1, w_2, \cdots, w_{k-1})\, t\right] - (1 - kp)\right\} dt > 0,$$

where

$$h(w_1, w_2, \cdots, w_{k-1}) = \frac{(A - B)w_1 + B \sum_{\alpha=1}^{k-1} w_\alpha}{\sqrt{n' + A \sum_{\alpha=1}^{k-1} w_\alpha^2 + B \sum_{\alpha \neq \beta} w_\alpha w_\beta}}.$$

The integrand on the left hand side is for all t a monotonically increasing function of $h(w_1, \cdots, w_{k-1})$, so the region where $pf_1 > (1 - kp)f_0$ must be of the type

$h(w_1, \cdots, w_{k-1}) > L$ where L is a number which depends on Δ/σ and p. The other regions can be calculated explicitly in a similar manner, and the Bayes solution is the following procedure: for $1 \leqq j \leqq k - 1$ select $\bar{D}_j$ if $w_j > 0$ and $w_j > \max(w_1, \cdots, w_{j-1}, w_{j+1}, \cdots, w_{k-1})$ and

$$(A - B)w_j + B \sum_{\alpha=1}^{k-1} w_\alpha > L \sqrt{n' + A \sum_{\alpha=1}^{k-1} w_\alpha^2 + B \sum_{\alpha \neq \beta} w_\alpha w_\beta};$$

select $\bar{D}_k$ if $w_j < 0$ for $j = 1, 2, \cdots, k - 1$ and

$$[-A - B(k-2)] \sum_{\alpha=1}^{k-1} w_\alpha > L \sqrt{n' + A \sum w_\alpha^2 + B \sum w_\alpha w_\beta};$$

otherwise select $\bar{D}_0$. Define the function $F(p)$ by the equation

$$F(p) = \int_0^\infty t^{n'+k-2} \exp\left(-\frac{t^2}{2}\right)\left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left(\frac{\Delta}{\sigma}\lambda_\alpha t\right) - (1 - kp)\right\} dt,$$

where λ_α is the constant used in (1). It is obvious that $F(p)$ is a continuous function of p with $F(0) < 0$ and $F(1/k) > 0$. Hence there exists a value p^* with $0 < p^* < 1/k$ which is a function of Δ/σ so that $F(p^*) = 0$. Once the Bayes solution relative to $(1 - kp, p, p, \cdots, p)$ has been worked out, it is obvious that to get the Bayes solution relative to $(1 - kp^*, p^*, \cdots, p^*)$ it is only necessary to replace L by λ_α. If we now substitute $w_i = (x_i - x_k)/s$ and replace A and B by their values, we find after some algebraic simplifications that the Bayes solution relative to $(1 - kp^*, p^*, \cdots, p^*)$ reduces to (1) when $\bar{D}_i$ is made to correspond to D_i. Since (1) is an allowable procedure, this proves that it is an optimum one.

3. The calculation of λ_α. The calculation of the exact value of λ_α required in order to have $P\{n(\bar{x}_M - \bar{x}) > \lambda_\alpha \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}\} = \alpha$ when all k means are equal will be extremely difficult until tables are made available, and therefore some approximation is required at present. For this purpose let A_i denote the event $[n(\bar{x}_i - \bar{x}) > \lambda_\alpha \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}]$, so that

$$P\{n(\bar{x}_M - \bar{x}) > \lambda_\alpha \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}\}$$

will be equal to the probability of the occurrence of at least one A_i $(i = 1, 2, \cdots, k)$. The approximation to be suggested is of a familiar type, and consists in determining λ_α so that $P(A_1) = \alpha/k$. For this purpose, it is clearly legitimate to take $m_1 = m_2 \cdots = m_k = 0$ and $\sigma = 1$. Next, let $y_j = \sqrt{n}\bar{x}_j$ $(j = 1, 2, \cdots, k)$ so that $\{y_j\}$ constitute a set of independent and standardized normal variables, and let $\bar{y} = (\sum_{i=1}^k y_i/k)$.. Then

$$P(A_1) = P\left\{y_1 - \bar{y} > \frac{\lambda_\alpha}{\sqrt{n}} \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^k (y_i - \bar{y})^2}\right\}.$$

Now we introduce an orthogonal transformation given by

$$t_1 = \frac{\sum_{i=1}^{k} y_i}{\sqrt{k}},$$

$$t_r = \frac{\sum_{i=r}^{k} y_i - (k - r + 1)y_{r-1}}{\sqrt{(k - r + 1)(k - r + 2)}}, \qquad r = 2, 3, \cdots, k.$$

The new variables $t_1, \cdots, t_k$ also constitute an independent set of normally distributed random variables with zero means and unit variances and are obviously independent of $\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$. We now have

$$\begin{aligned} P(A_1) &= P\left\{\sqrt{\frac{k-1}{k}}\,(-t_2) > \frac{\lambda_\alpha}{\sqrt{n}} \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=2}^{k} t_i^2}\right\} \\ &= \tfrac{1}{2}P\left\{\frac{(k-1)}{k}\, t_2^2 > \frac{\lambda_\alpha^2}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=2}^{k} t_i^2\right)\right\} \\ &= \tfrac{1}{2}P\left\{\left(\frac{k-1}{k} - \frac{\lambda_\alpha^2}{n}\right) t_2^2 > \frac{\lambda_\alpha^2}{n}\chi_{n''}^2\right\}, \end{aligned}$$

where $n'' = k(n - 1) + k - 2$ and $\chi_{n''}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=3}^{k} t_i^2$ has the chi-square distribution with n'' degrees of freedom and is independent of t_2. If F_0 is used for the value of the F distribution with $n_1 = 1$ and $n_2 = n''$ degrees of freedom which is exceeded with probability $2\alpha/k$, it is a simple matter to verify that the desired approximation is given by

$$\lambda_\alpha = \sqrt{\frac{n(k-1)F_0}{k(n'' + F_0)}}.$$

If λ_α is determined by the above formula so that $P(A_1) = \alpha/k$, it follows at once from Bonferoni's inequality [5] that the probability of not selecting D_0 when all the means are equal will be less than α by an amount which cannot exceed $\frac{1}{2}k(k - 1)P(A_1A_2)$. This quantity is still difficult to evaluate, but in the limit as $n \to \infty$, $\frac{1}{2}k(k - 1)P(A_1A_2)$ can be obtained from tables of the normal bivariate distribution, and is easily shown to be less than $\frac{1}{2}\alpha^2$ for n large enough. Even for small n it seems plausible on an intuitive basis that this bound will be small for values of α ordinarily of interest (say $\alpha \leqq .05$), although further investigation on this point would obviously be desirable. In any event, if the approximation $\lambda_\alpha = \sqrt{n(k - 1)F_0/[k(n'' + F_0)]}$ is used, it can be asserted that for any n the probability of not selecting D_0 when all the means are equal is less than α, and for large n the difference between the true probability and α will be less than $\frac{1}{2}\alpha^2$.

REFERENCES

[1] Frederick Mosteller, "A k-sample slippage test for an extreme population," *Annals of Math. Stat.*, Vol 19 (1948), pp. 58–65.

[2] Edward Paulson, "A multiple decision procedure for certain problems in the analysis of variance," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 95–98.

[3] Edward Paulson, "The comparison of several experimental categories with a control," *Annals of Math. Stat.*, Vol. 23 (June 1952), pp. 239–246.

[4] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[5] William Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, 1950.

LIMIT THEOREMS ASSOCIATED WITH VARIANTS OF THE VON MISES STATISTIC[1]

By M. Rosenblatt

University of Chicago

1. Summary. A multidimensional analogue of the von Mises statistic is considered for the case of sampling from a multidimensional uniform distribution. The limiting distribution of the statistic is shown to be that of a weighted sum of independent chi-square random variables with one degree of freedom. The weights are the eigenvalues of a positive definite symmetric function.

A modified statistic of the von Mises type useful in setting up a two sample test is shown to have the same limiting distribution under the null hypothesis (both samples come from the same population with a continuous distribution function) as that of the one-dimensional von Mises statistic. We call the statistics mentioned above von Mises statistics because they are modifications of the ω^2 criterion considered by von Mises [5].

The paper makes use of elements of the theory of stochastic processes.

2. Introduction. Let $X_i = (X_{i1}, \cdots, X_{ik})$, $i = 1, \cdots, n$, be a sample from a k-dimensional uniform distribution; that is, x_{ij}, $i = 1, \cdots, n, j = 1, \cdots, k$, are independent and uniformly distributed on [0, 1]. Let

$$\phi_t(x) = \begin{cases} 1 \text{ if } x \leqq t, \\ 0 \text{ if } x > t. \end{cases} \tag{1}$$

The sample distribution function is

$$S_n(\bar{t}) = S_n(t_1, \cdots, t_k) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{k} \phi_{t_j}(X_{ij}), \tag{2}$$

where $\bar{t} = (t_1, \cdots, t_k)$. Consider the process

$$Y_n(\bar{t}) = \sqrt{n}(S_n(t_1, \cdots, t_k) - t_1 \cdots t_k), \qquad 0 \leqq t_1, \cdots, t_k \leqq 1. \tag{3}$$

Clearly $EY_n(\bar{t}) = 0$. The covariance of the process is

$$\begin{aligned} E(Y_n(\bar{t})Y_n(\bar{t}')) &= r_n(\bar{t}, \bar{t}') \\ &= \frac{1}{n} E\left[\sum_{i=1}^{n}\left\{\prod_{j=1}^{k} \phi_{t_j}(X_{ij}) - t_1 \cdots t_k\right\} \sum_{i=1}^{n}\left\{\prod_{j=1}^{k} \phi_{t'_j}(X_{ij}) - t'_1 \cdots t'_k\right\}\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} E\left(\left\{\prod_{j=1}^{k} \phi_{t_j}(X_{ij}) - t_1 \cdots t_k\right\}\left\{\prod_{j=1}^{k} \phi_{t'_j}(X_{ij}) - t'_1 \cdots t'_k\right\}\right) \\ &= \prod_{j=1}^{k} \min(t_j, t'_j) - t_1 \cdots t_k t'_1 \cdots t'_k. \end{aligned} \tag{4}$$

---

[1] Work done under contract with the Office of Naval Research.

Note that the covariance function $r_n(\bar{t}, \bar{t}')$ is independent of n and symmetric in $\bar{t}$ and $\bar{t}'$.

Consider the function $r_n(\bar{t}, \bar{t}') = r(\bar{t}, \bar{t}')$ as the kernel in the following eigenvalue problem

$$\int_0^1 r(\bar{t}, \bar{t}')\phi(\bar{t}')\, d\bar{t}' = \lambda\phi(\bar{t}),$$

where the integral is over all components of $\bar{t}'$. The kernel is positive definite (being a covariance function) and hence all its eigenvalues are positive. There are a denumerable number of eigenvalues. Denote the eigenvalues by $\lambda_1, \lambda_2, \cdots$ and the corresponding orthonormal eigenfunctions by

$$\phi_1(\bar{t}), \qquad \phi_2(\bar{t}), \qquad \cdots .$$

It is understood that each eigenvalue is repeated with the multiplicity of the linearly independent eigenfunctions corresponding to it. Now

$$r(\bar{t}, \bar{t}') = \sum_{j=1}^{\infty} \lambda_j \phi_j(\bar{t})\phi_j(\bar{t}') \tag{5}$$

with uniform convergence according to Mercer's theorem. The general theorem of Karhunen on representation of stochastic processes [3] then implies that

$$Y_n(\bar{t}) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, \phi_j(\bar{t}) Y_{nj} \tag{6}$$

in the mean square, where

$$EY_{nj} = 0, \qquad EY_{nj}Y_{nk} = \delta_{jk} .$$

3. The limiting distribution. As $n \to \infty$, the joint distribution of $Y_n(\bar{t}_1), \cdots, Y_n(\bar{t}_m)$ approaches the joint distribution of $Y(\bar{t}_1), \cdots, Y(\bar{t}_m)$, where $Y(\bar{t})$ is a normal process with mean zero and covariance $r(\bar{t}, \bar{t}')$. Obviously the process

$$Y(\bar{t}) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, \phi_j(\bar{t}) Y_j , \tag{7}$$

where the Y_j are independent normal random variables with mean zero and variance one.

THEOREM 1. *The von Mises statistic corresponding to a sample of n from a k-dimensional uniform distribution is*

$$\int_0^1 Y_n^2(\bar{t})\, d\bar{t} = \sum_{j=1}^{\infty} \lambda_j Y_{nj}^2 , \tag{8}$$

and the limiting distribution of (8) *as* $n \to \infty$ *is that of*

$$\int_0^1 Y^2(\bar{t})\, d\bar{t} = \sum_{j=1}^{\infty} \lambda_j Y_j^2 . \tag{9}$$

PROOF. Now

$$Y_n(\bar{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(\bar{t}),$$

where the random variables

$$Z_i(\bar{t}) = \prod_{j=1}^{k} \phi_{t_j}(X_{ij}) - t_1 \cdots t_k$$

are independent and identically distributed. Then

$$Y_{nj} = \sum_{k=1}^{n} \frac{1}{\sqrt{n}} Z_{kj},$$

where

$$Z_{kj} = \frac{1}{\sqrt{\lambda_j}} \int_0^1 Z_k(\bar{t}) \phi_j(\bar{t})\, d\bar{t}.$$

The random vectors

$$(Z_{k1}, \cdots, Z_{kN}), \qquad k = 1, \cdots, n,$$

are independent and identically distributed. Moreover

$$EZ_{kj} = 0,$$

$$EZ_{kj}Z_{kl} = \delta_{jl}.$$

The multidimensional central limit theorem then implies that the random variables Y_{nj}, $j = 1, \cdots, N$, are asymptotically normal, independent random variables with mean zero and variance one as $n \to \infty$. $Y_{n1}^2, \cdots, Y_{nN}^2$ as $n \to \infty$ are asymptotically independent chi-square random variables with one degree of freedom and mean one. Note that

$$\int_0^1 r(\bar{t}, \bar{t})\, d\bar{t} = \sum_{j=1}^{\infty} \lambda_j. \tag{10}$$

Given any $\epsilon > 0$, let $N(\epsilon)$ be such that

$$\sum_{N(\epsilon)+1}^{\infty} \lambda_j < \epsilon^2.$$

Let

$$U_N = \sum_{j=1}^{N(\epsilon)} \lambda_j Y_{nj}^2,$$

$$V_N = \sum_{N(\epsilon)+1}^{\infty} \lambda_j Y_{nj}^2.$$

U_N asymptotically has the same distribution as

$$\sum_{1}^{N(\epsilon)} \lambda_j Y_j^2;$$

that is, for sufficiently large n

$$\left| P\{U_N \leqq x\} - P\left\{\sum_1^{N(\epsilon)} \lambda_j Y_j^2 \leqq x\right\}\right| < \epsilon.$$

The choice of $N(\epsilon)$ and Tchebycheff's inequality imply

$$\left| P\left\{\int_0^1 Y^2(\bar{t})\, d\bar{t} \leqq x + \epsilon\right\} - P\left\{\sum_1^{N(\epsilon)} \lambda_j Y_j^2 \leqq x\right\}\right| < \epsilon,$$

$$\left| P\{U_N \leqq x\} - P\left\{\int_0^1 Y_N^2(\bar{t})\, d\bar{t} \leqq x + \epsilon\right\}\right| < \epsilon.$$

Hence

$$\left| P\left\{\int_0^1 Y_n^2(\bar{t})\, d\bar{t} \leqq x + \epsilon\right\} - P\left\{\int_0^1 Y^2(\bar{t})\, d\bar{t} \leqq x + \epsilon\right\}\right| < 3\epsilon$$

for sufficiently large n. The distribution function of $\int_0^1 Y^2(\bar{t})\, d\bar{t}$ is continuous. Therefore the limiting distribution of $\int_0^1 Y_n^2(\bar{t})\, d\bar{t}$ as $n \to \infty$ is the same as that of $\int_0^1 Y^2(\bar{t})\, d\bar{t}$.

The distribution function of $\int_0^1 Y^2(\bar{t})\, d\bar{t}$ has been computed in the 1-dimensional case ($k = 1$). The eigenvalues of (4) are then $\lambda_j = 1/(\pi^2 j^2)$ $j = 1, 2, \cdots$ and hence the characteristic function of (9) is $\prod_{j=1}^{\infty} [1 - 2it/(\pi^2 j^2)]^{-\frac{1}{2}}$. One can invert the characteristic function by a contour integration and obtain the distribution function of (9) as given by Smirnov [5, 2]. It would be of great interest to find the eigenvalues of (4) when $k > 1$.

4. The two sample test. Let X_{1j}, $j = 1, \cdots, n$, and X_{2k}, $k = 1, \cdots, m$, be samples of n and m respectively from a population with some continuous distribution function $F(x)$. Let $S_1(t)$, $S_2(t)$ be the corresponding sample distribution functions. Various people [4] have suggested using

$$\frac{mn}{m+n} \int_{-\infty}^{\infty} (S_1(t) - S_2(t))^2\, d\left(\frac{S_1(t) + S_2(t)}{2}\right) \tag{11}$$

as a test statistic for the two sample problem.

THEOREM 2. *Statistic* (11) *has the same limiting distribution when* $n \to \infty$, $m/n \to \lambda > 0$ *as the one-dimensional von Mises statistic under the assumption that both samples come from the same continuous population.*

Consider computing the statistic for samples $F(X_{1j})$, $j = 1, \cdots, n$, $F(X_{2k})$, $k = 1, \cdots, m$, of n and m respectively from a population with the uniform distribution. The value of the statistic is the same as that obtained from the original samples $\{X_{1j}\}$, $\{X_{2k}\}$ and consequently has the same distribution function as the latter. We need then only consider the statistic

$$\frac{mn}{m+n}\int_0^1 (S_1(t) - S_2(t))^2\, d\left(\frac{S_1(t)+S_2(t)}{2}\right) \tag{12}$$

when the samples are from a uniformly distributed population. It is obvious that

$$\frac{mn}{m+n}\int_0^1 (S_1(t) - S_2(t))^2\, dt \tag{13}$$

has the same limiting distribution as the one-dimensional von Mises statistic when $n \to \infty$, $m/n \to \lambda > 0$. It would then be sufficient to show that

$$\frac{mn}{m+n}\int_0^1 (S_1(t) - S_2(t))^2\, d\left(\frac{S_1(t)+S_2(t)}{2} - t\right) \tag{14}$$

converges to zero in probability when $n \to \infty$, $m/n \to \lambda > 0$. Now

$$\begin{aligned}
&\frac{mn}{m+n}\int_0^1 (S_1(t) - S_2(t))^2\, d\left(\frac{S_1(t)+S_2(t)}{2} - t\right)\\
&\quad= \frac{mn}{m+n}\int_0^1 (S_1(t) - S_2(t))^2\, d(S_2(t) - t)\\
&\quad= \frac{mn}{m+n}\int_0^1 (S_1(t) - t)^2\, d(S_2(t) - t) + \frac{mn}{m+n}\int_0^1 (S_2(t) - t)^2\, d(S_1(t) - t)
\end{aligned}$$

can be obtained by a series of integrations by parts. The proof is complete if one can show that both expressions directly above converge to zero in probability. By symmetry it is enough to consider one of the expressions.

Let

$$\begin{aligned} x_1(t) &= n^{\frac{1}{2}}(S_1(t) - t),\\ x_2(t) &= m^{\frac{1}{2}}(S_2(t) - t). \end{aligned} \tag{15}$$

Now

$$\begin{aligned} Ex_i(t) &= 0,\\ Ex_i(\tau)x_i(t) &= \min(\tau, t) - \tau t, \qquad i = 1, 2. \end{aligned} \tag{16}$$

We use the following transformation suggested by Doob [1]

$$x_i(t) = (t-1)Z_i\left(\frac{t}{1-t}\right), \qquad i = 1, 2. \tag{17}$$

The processes $Z_1(t)$, $Z_2(t)$ are independent of each other. Moreover, each of them is an orthogonal process with

$$\begin{aligned} EZ_i(t) &= 0,\\ EZ_i(t)Z_i(\tau) &= \min(\tau, t), \qquad i = 1, 2. \end{aligned} \tag{18}$$

A simple computation making use of (2), (15), (17) yields

$$EZ_1^2(t)Z_1^2(\tau) = \frac{1}{n}\left[\frac{\min(t,\tau)}{1+\min(t,\tau)} + \frac{\min(t,\tau)(\max(t,\tau) - \min(t,\tau))}{(1+t)(1+\tau)} + \frac{t^2\tau^2}{1+\max(\tau,t)}\right] + \frac{n-1}{n} t\tau + 2\frac{n-1}{n}(\min(t,\tau))^2 \tag{19}$$

and in particular

$$EZ_1^4(t) = \frac{1}{n}\frac{t+t^4}{1+t} + 3\frac{n-1}{n}t^2. \tag{20}$$

Now

$$\frac{mn}{m+n}\int_0^1 (S_1(t) - t)^2\, d(S_2(t) - t) = \frac{m^{\frac{1}{2}}}{m+n}\int_0^1 x_1^2(t)\, dx_2(t)$$

$$= \frac{m^{\frac{1}{2}}}{m+n}\int_0^1 (t-1)^2 Z_1^2\left(\frac{t}{1-t}\right) Z_2\left(\frac{t}{1-t}\right) dt + \frac{m^{\frac{1}{2}}}{m+n}\int_0^1 (t-1)^3 Z_1^2\left(\frac{t}{1-t}\right) dZ_2\left(\frac{t}{1-t}\right) \tag{21}$$

$$= \frac{m^{\frac{1}{2}}}{m+n}\int_0^\infty \frac{1}{(t+1)^4} Z_1^2(t)Z_2(t)\, dt \tag{22}$$

$$+ \frac{m^{\frac{1}{2}}}{m+n}\int_0^\infty \frac{1}{(t+1)^3} Z_1^2(t)\, dZ_2(t). \tag{23}$$

The random variables (22), (23) are the limits almost everywhere of

$$\frac{m^{\frac{1}{2}}}{m+n}\int_0^T \frac{1}{(t+1)^4} Z_1^2(t)Z_2(t)\, dt \tag{24}$$

and

$$\frac{m^{\frac{1}{2}}}{m+n}\int_0^T \frac{1}{(t+1)^3} Z_1^2(t)\, dZ_2(t), \tag{25}$$

respectively, as $T \to \infty$. The independence of the orthogonal processes $Z_1(t)$, $Z_2(t)$ implies that the second moments of (24), (25) are

$$\frac{m}{(m+n)^2}\int_0^T\int_0^T \frac{\min(\tau,t)E(Z_1^2(t)Z_1^2(\tau))}{(1+t)^4(1+\tau)^4}\, dt\, d\tau$$

and

$$\frac{m}{(m+n)^2}\int_0^T \frac{1}{(1+t)^6} E(Z_1^4(t))\, dt,$$

respectively.

Making use of (19), (20) one can see that (24), (25) converge in mean square as $T \to \infty$ to (22), (23) respectively. But then the second moments of (22), (23) exist and are given by

$$\frac{m}{(m+n)^2} \int_0^\infty \int_0^\infty \frac{\min(\tau, t) E(Z_1^2(t) Z_1^2(\tau))}{(1+t)^4 (1+\tau)^4} \, dt \, d\tau \tag{26}$$

and

$$\frac{m}{(m+n)^2} \int_0^\infty \frac{1}{(1+t)^6} E(Z_1^4(t)) \, dt, \tag{27}$$

respectively. The second moments (26), (27) converge to zero as $n \to \infty$, $m/n \to \lambda > 0$ and hence the random variables (22), (23) converge to zero in probability as $n \to \infty$, $m/n \to \lambda > 0$. This in turn implies that (21) converges to zero in probability. The same argument implies that

$$\frac{mn}{m+n} \int_0^1 (s_2(t) - t)^2 \, d(s_1(t) - t)$$

converges to zero in probability. Hence (14) converges to zero in probability as $n \to \infty$, $m/n \to \lambda > 0$ and the proof is complete.

REFERENCES

[1] J. L. Doob, "Heuristic approach to Kolmogorov-Smirnov theorems," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[2] M. Kac, "On some connections between probability theory and differential and integral equations," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 1951.

[3] K. Karhunen, "Uber lineare Methoden in der Warscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae*, Series *A. I.*, no. 37 (1947), 79 pp.

[4] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[5] N. Smirnoff, "Sur la distribution de ω^2," *Comptes Rendus de l'Academie des Sciences*, Vol. 202 (1936), p. 449.

NOTES

A MARKOV CHAIN DERIVATION OF DISCRETE DISTRIBUTIONS

BY F. G. FOSTER

Magdalen College, Oxford

Let an irreducible, aperiodic Markov chain[1] have the matrix of transition probabilities, $\mathbf{A} = [p_{ij}]$ $(i, j = 0, 1, 2, \cdots)$. Then as usual we shall have

$$p_{ij} \geqq 0 \qquad \text{for all } i \text{ and } j,$$

$$\sum_{j=0}^{\infty} p_{ij} = 1 \qquad \text{for all } i.$$

It is known ([1], p. 325) that the nth power of $\mathbf{A}$, $\mathbf{A}^n$, tends to a limiting matrix as $n \to \infty$

$$\lim_{n\to\infty} \mathbf{A}^n = \mathbf{B},$$

and $\mathbf{B}$ will either be null or have the identical rows,

$$\mathbf{x} = (x_0, x_1, \cdots),$$

such that $x_i > 0$ for all i and $\sum_{i=0}^{\infty} x_i = 1$. Moreover we shall have

$$\mathbf{xA} = \mathbf{x}.$$

In this way we may make correspond to any matrix $\mathbf{A}$, of the type under consideration, either the null vector or a probability distribution represented by $\mathbf{x}$. Conversely, to any distribution $\mathbf{x}$ there will correspond a matrix $\mathbf{A}$ (not necessarily unique). A method of constructing such a matrix is given below and illustrated with some examples.

Let $\{a_i\}$ $(i = 0, 1, 2, \cdots)$ be a sequence of positive numbers and define $A_n \doteq \sum_{i=0}^{n} a_i$ $(n = 0, 1, 2, \cdots)$. Now let

$$\mathbf{A} = \begin{bmatrix} \frac{a_0}{A_1} & \frac{a_1}{A_1} & 0 & 0 & 0 & \cdots \\ \frac{a_0}{A_2} & \frac{a_1}{A_2} & \frac{a_2}{A_2} & 0 & 0 & \cdots \\ \frac{a_0}{A_3} & \frac{a_1}{A_3} & \frac{a_2}{A_3} & \frac{a_3}{A_3} & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Then $\mathbf{A}$ satisfies the usual conditions for being a transition probability matrix;

[1] For definitions of all terms used see [1].

moreover it is clearly irreducible, and, since its diagonal elements are all positive, it is also aperiodic. Now suppose $\mathbf{x}$ is a vector such that

$$\mathbf{xA} = \mathbf{x}.$$

If we regard this as an equation in $\mathbf{x}$, we find that in our special case the solution is easily obtained. We have

$$x_1 = \frac{a_1}{a_0} x_0,$$

$$\frac{x_{n+1}}{a_{n+1}} = \frac{x_n}{a_n} - \frac{x_{n-1}}{A_n}, \qquad n \geqq 1.$$

It follows by induction that

$$x_n = x_0 \frac{a_1 a_2 \cdots a_n}{A_0 A_1 \cdots A_{n-1}}, \qquad n \geqq 1,$$

and so $\mathbf{x}$ is uniquely determined, apart from a common factor. For $\mathbf{x}$ to be a distribution, we must have in addition

$$1 = \sum_{n=0}^{\infty} x_n = x_0 + x_0 \sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}}.$$

Thus

$$x_0 = 1 \Big/ \left(1 + \sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}}\right),$$

and it follows that the matrix $\mathbf{B}$ ($= \lim \mathbf{A}^n$) is non-null if and only if

$$\sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} < \infty,$$

and each row of $\mathbf{B}$ will consist of the distribution $\mathbf{x}$.

Conversely, if we have given a distribution $\mathbf{x}$, we may calculate the sequence $\{a_i\}$ which possesses the required property. We have only to put

$$\frac{a_n}{A_{n-1}} = \frac{x_n}{x_{n-1}}, \qquad n \geqq 1.$$

Then we find as required that

$$x_0 \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} = x_n.$$

The sequence $\{a_i\}$ is now easily calculated. We have

$$A_n - A_{n-1} = a_n = A_{n-1} \frac{x_n}{x_{n-1}}.$$

Therefore

$$A_n = A_{n-1}\left(1 \quad \frac{x_n}{x_{n-1}}\right),$$

and, by iteration,

$$A_n = a_0 \prod_{i=1}^{n} \left(1 + \frac{x_i}{x_{i-1}}\right).$$

Hence, putting (as we may, since a common factor is unimportant) $a_0 = 1$, we have

$$a_n = \frac{x_n}{x_{n-1}} \prod_{i=1}^{n-1} \left(1 + \frac{x_i}{x_{i-1}}\right), \qquad n \geqq 1.$$

The above procedure may be given the following interpretation. Consider a particle performing a random walk on the integers 0, 1, 2, $\cdots$ in such a manner that when in position n it has probabilities in the ratios

$$a_0 : a_1 : a_2 : \cdots : a_{n+1}$$

of jumping at the next move into one of the positions 0, 1, 2, $\cdots$, $n + 1$. That is, the particle can move either one step along or back to any previous position. The distribution $\{x_i\}$ may then be interpreted as giving the asymptotic probabilities for its position after a large number of moves, and we have shown how the sequence $\{a_i\}$ may be calculated to give any required asymptotic distribution $\{x_i\}$ (with $x_i > 0$ for all i).

In some cases where $\mathbf{x}$ is a recognised distribution the sequence $\{a_i\}$ has a particularly simple form.

Example (a). *The Poisson distribution.* Let us take

$$x_n = e^{-\lambda}\lambda^n/n!, \qquad \lambda > 0, \qquad n = 0, 1, 2 \cdots .$$

We find that

$$a_n = \frac{1}{n!}\lambda(\lambda + 1) \cdots (\lambda + n - 1), \qquad n = 1, 2, \cdots ,$$

with $a_0 = 1$. Thus the nth row of $\mathbf{A}$ is a truncated negative binomial distribution having $n + 1$ terms. In particular, when $\lambda = 1$, $\mathbf{A}$ takes the very simple form wherein $a_n \equiv 1$ for all n, and the rule governing the motion of the particle is that when it is in the nth position it has *equal* probabilities of jumping into any of the positions 0, 1, 2, $\cdots$, $n + 1$. We have then the result that its asymptotic position is a random variable with the Poisson distribution $\{1/(en!)\}$ ($n = 0, 1, 2, \cdots$).

Example (b). *The negative binomial distribution.* Let us take

$$x_n = (1 - \beta)^\lambda \frac{1}{n!}\lambda(\lambda + 1) \cdots (\lambda + n - 1)\beta^n, \quad n = 1, 2, \cdots ,$$

$$x_0 = (1 - \beta)^\lambda,$$

where $0 < \beta < 1, \lambda > 0$. We find that

$$a_n = \frac{\lambda + n - 1}{n!} \beta(1 + \lambda\beta) \cdots (n - 1 + [\lambda + n - 2]\beta), \qquad n = 1, 2, \cdots,$$

with $a_0 = 1$. In particular, when $\lambda = 1$, we have

$$a_n = (1 + \beta)^{n-1}\beta, \qquad n = 1, 2, \cdots,$$

with $a_0 = 1$, and each row of **A** is a truncated modified geometric distribution.

REFERENCE

[1] WILLIAM FELLER. *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, 1950.

ON MINIMUM VARIANCE ESTIMATORS[1]

BY J. KIEFER

Cornell University

Chapman and Robbins [1] have given a simple improvement on the Cramér-Rao inequality without postulating the regularity assumptions under which the latter is usually proved. The purpose of this note is to show by examples how a similarly derived stronger inequality (see equation (2)) may be used to verify that certain estimators are uniform minimum variance unbiased estimators. This stronger inequality is that which (under additional restrictions) was shown in [2] to be the best possible, but is in a more useful form for applications than the form given in [2]. For simplicity we consider only an inequality on the variance of unbiased estimators, but inequalities on other moments than the second (see [2]), or for biased estimators, may be found similarly. The two examples considered here are ones where the regularity conditions of [2] are not satisfied, where the method of [1] does not give the best bound, and where the method of this note is used to find the best bound and thus to verify that certain estimators are uniform minimum variance unbiased. (For the examples considered this also follows from completeness of the sufficient statistic; the method used here applies, of course, more generally.)

Let X be a chance variable with density $f(x; \theta)$ with respect to some fixed σ-finite measure μ. ($\theta \in \Omega$, $x \in \mathfrak{X}$). We suppose suitable Borel fields to be given and $f(x; \theta)$ to be measurable in its arguments. Ω is a subset of the real line. For each θ, let $\Omega_\theta = \{h \mid (\theta + h) \in \Omega\}$. For fixed θ, let λ_1 and λ_2 be any two probability measures on Ω_θ such that $E_i h = \int_{\Omega_\theta} h d\lambda_i(h)$ exists for $i = 1, 2$. Then, for any

[1] Research sponsored by the Office of Naval Research.

$t(x)$ for which $E_\theta t = \theta$, we have

$$(1)\quad \int_{\mathfrak{X}} (t-\theta)\sqrt{f(x;\theta)}\left\{\frac{\int_{\Omega_\theta} f(x;\theta+h)\,d[\lambda_1(h)-\lambda_2(h)]}{f(x;\theta)}\right\}\sqrt{f(x;\theta)}\,d\mu = E_1h - E_2h.$$

Applying Schwarz's inequality, we have after some obvious manipulations,

$$(2)\quad E_\theta(t-\theta)^2 \geqq \sup\left\{\frac{(E_1h - E_2h)^2}{\int_{\mathfrak{X}} \frac{\left\{\int_{\Omega_\theta} f(x;\theta+h)\,d[\lambda_1(h)-\lambda_2(h)]\right\}^2}{f(x;\theta)}\,d\mu}\right\},$$

where for each θ the supremum is taken over all λ_1 and λ_2 for which $\lambda_1 \neq \lambda_2$ and for which the integrand of the integral over $\mathfrak{X}$ is defined a.e. (μ).

We remark that the supremum of (2) is easily seen to be unimproved if λ_i and E_ih are multiplied by real numbers c_i $(i = 1, 2)$ with respect to which the supremum is also taken. From this fact it is easy to verify that the right side of (2) must coincide with the expression given in Theorem 4 of [2] (for $s = 2$ there), and which Barankin shows (under the assumption that $f(x;\theta+h)/f(x;\theta)$ is defined a.e. (μ) and (for our case) belongs to L_2 with respect to the measure $\nu(A) = \int_A f(x;\theta)\,d\mu$ for all $h \in \Omega_\theta$) to be the best possible bound. However, the form of equation (2) is more useful for applications, since one can sometimes find λ_i for which the bound is attained but where no discrete λ_i (essentially what are used in the form of [2]) actually give this bound.

It will often suffice in applications to let λ_2 give measure one to the point $h = 0$. This gives

$$(3)\quad E_\theta(t-\theta)^2 \geqq \sup_{\lambda_1}\left\{\frac{(E_1h)^2}{\int_{\mathfrak{X}} \frac{\left[\int_{\Omega_\theta} f(x;\theta+h)\,d\lambda_1(h)\right]^2}{f(x;\theta)}\,d\mu - 1}\right\}.$$

If we consider only those λ_1 which give measure one to a single h, we obtain

$$(4)\quad E_\theta(t-\theta)^2 \geqq \frac{1}{\inf_h \frac{1}{h^2}\left\{\int_{\mathfrak{X}} \frac{[f(x;\theta+h)]^2}{f(x;\theta)}\,d\mu - 1\right\}},$$

where the infimum is over all $h \neq 0$ for which $h \in \Omega_\theta$ and for which $f(x;\theta) = 0$ implies $f(x;\theta+h) = 0$ a.e. (μ). The latter is precisely the condition of equation (2) of [1], the result of which thus coincides with (4).

We now give two examples where the right side of (3) suffices to give the best bound, where the right side of (4) does not give the best bound, and where the previously mentioned restrictions of [2] are not satisfied. In both examples μ is Lebesgue measure on the real line.

EXAMPLE 1. We have n observations from a rectangular distribution from 0 to θ ($\Omega = \{\theta \mid \theta > 0\}$). It suffices to consider the maximum Y of the observations, whose density is ny^{n-1}/θ^n for $0 \leqq y \leqq \theta$, and 0 elsewhere. For $n = 1$, the denominator of the right side of (4) becomes $\inf_{-\theta<h<0}\{-1/[h(\theta + h])\}$, so that (4) gives the bound $\theta^2/4$. It would be too tedious to carry this calculation out for each n, but it can be shown that, as $n \to \infty$, (4) asymptotically gives the bound $.648\theta^2/n^2$. On the other hand, if we put $d\lambda_1(h) = [(n + 1)/\theta]\,(h/\theta + 1)^n\,dh$ for $-\theta < h < 0$, the term in braces on the right side of (3) becomes $\theta^2/[n(n + 2)]$, which is in fact attained as the variance of the unbiased estimator $[(n + 1)/n]Y$.

EXAMPLE 2. We have m observations from the distribution with density $e^{-(x-\theta)}$ for $x \geqq \theta$ and 0 elsewhere (Ω is the real line). Here the minimum Z of the observations is sufficient and has density $me^{-m(z-\theta)}$, $z \geqq \theta$. The denominator of (4) is $\inf_{h>0}\,([e^{mh} - 1]/h^2)$. The infimum is attained for $mh = 1.5936$, and yields $.648/m^2$ as the bound given by (4). On the other hand, putting $d\lambda_1(h) = me^{-mh}\,dh$ for $0 < h < \infty$ and 0 otherwise, the expression in braces of (3) becomes $1/m^2$, which is actually attained as the variance of the unbiased estimator $Z - 1/m$.

REFERENCES

[1] D. C. CHAPMAN AND H. ROBBINS, "Minimum variance estimation without regularity assumptions," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 581–586.

[2] E. BARANKIN, "Locally best unbiased estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501.

BHATTACHARYYA BOUNDS WITHOUT REGULARITY ASSUMPTIONS

BY D. A. S. FRASER AND IRWIN GUTTMAN

University of Toronto

1. Summary. In [1] a method for removing the regularity conditions from the Cramér-Rao Inequality was given and applied to the estimation of a single real parameter. It was noted there that the method would extend to problems more general than estimating a single real parameter. However, the method extends also for the estimation of a single real parameter and produces analogues of the Bhattacharyya bounds with and without nuisance parameters.

2. Introduction. Let $\mu(x)$ be a σ-finite measure defined over an additive class $\mathfrak{A}$ of subsets of a space $\mathfrak{X}$, and let X be a random variable with density

$$f(x; \theta_1, \cdots, \theta_k)$$

with respect to $\mu(x)$. $\theta_1, \cdots, \theta_k$ are real with $(\theta_1, \cdots, \theta_k) = \Theta \,\varepsilon\, A \subset R^k$. The carrier $S(\theta_1, \cdots, \theta_k)$ of the distribution is defined by

$$S(\theta_1, \cdots, \theta_k) = \{x \mid f(x; \theta_1, \cdots, \theta_k) > 0\}.$$

We restrict our consideration of $S(\theta_1, \cdots, \theta_k)$ to the positive sample space of the measure μ.

The following lemma given in [2] will be needed:

LEMMA. *If the real valued random variables $T, S_1, \cdots, S_r$ satisfy*

$$
\begin{aligned}
E(S_i) &= 0, & & \\
E(TS_i) &= 1, & & i = 1, \\
&= 0, & & i = 2, \cdots, r,
\end{aligned}
$$

then $\sigma_T^2 \geqq 1/\sigma^2_{S_1 \cdot S_2 \cdots S_r}$, where $\sigma^2_{S_1 \cdot S_2 \cdots S_r}$ is the variance of the residual of the regression fit of $S_2, \cdots, S_r$ to S_1.

PROOF. Since the covariance of T and $S_1 - \sum_2^r l_i S_i$ is 1, then the product of the variances is greater than or equal to 1; $\sigma_T^2 \geqq 1/\sigma^2_{S_1 - \Sigma l_i S_i}$. The sharpest inequality is obtained by a regression fit.

3. Bhattacharyya bounds. Let T be an unbiased estimate of the parameter θ_1, and define $\{S_{i_1 \cdots i_k}\}$ as follows:

$$
\begin{aligned}
S_1 &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}}{\Delta} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \\
&= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \cdot \frac{f(x; \theta_1^{(1)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)}) - f(x; \theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{\theta_1^{(1)} - \theta_1^{(0)}}, \\
S_2 &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}, \theta_1^{(2)}}{\Delta^2} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \\
&= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(2)}}{\Delta} \underset{\theta_1^{(1)}}{\Delta} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \qquad (3.1) \\
&= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \left[\frac{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(0)} - \theta_1^{(1)})(\theta_1^{(0)} - \theta_1^{(2)})} \right. \\
&\qquad \left. + \frac{f(x; \theta_1^{(1)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(1)} - \theta_1^{(0)})(\theta_1^{(1)} - \theta_1^{(2)})} + \frac{f(x; \theta_1^{(2)}, \theta_2^{(0)}, \cdots, \theta_k^{(0)})}{(\theta_1^{(2)} - \theta_1^{(0)})(\theta_1^{(2)} - \theta_1^{(1)})} \right], \\
S_{i_1, \cdots, i_k} &= \frac{1}{f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)})} \underset{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}}{\Delta^{i_1}} \cdots \underset{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}}{\Delta^{i_k}} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}),
\end{aligned}
$$

where $\Delta^i_{\theta^{(1)} \cdots \theta^{(i)}} g(\theta^{(0)})$ is the ith divided difference

$$
\begin{aligned}
\underset{\theta^{(1)}, \cdots, \theta^{(i)}}{\Delta^i} g(\theta^{(0)}) &= \underset{\theta^{(i)}}{\Delta} \cdots \underset{\theta^{(1)}}{\Delta} g(\theta^{(0)}), \\
\underset{\theta^{(1)}}{\Delta} g(\theta^{(0)}) &= \frac{g(\theta^{(1)}) - g(\theta^{(0)})}{\theta^{(1)} - \theta^{(0)}},
\end{aligned}
$$

where the expressions are to be considered as functions of $\theta^{(0)}$ for further differencing. Also we introduce the following assumption concerning the carrier of the distribution.

Assumption A. *$S(\theta_1^{(0)}, \cdots, \theta_k^{(0)}) \supset S(\theta_1^{(i_1)}, \cdots, \theta_k^{(i_k)})$ for all $(i_1, \cdots, i_k)$ for which S's have been defined.*

On the basis of Assumption A, we have

$$E_{\Theta_0}(S_{i_1 \cdots i_k}) = \int \Delta^{i_1}_{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}} f(x; \theta_1^{(0)}, \cdots, \theta_k^{(0)}) \, d\mu(x)$$
$$= 0,$$

$$E_{\Theta_0}(TS_{i_1 \cdots i_k}) = \int \Delta^{i_1}_{\theta_1^{(1)}, \cdots, \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)}, \cdots, \theta_k^{(i_k)}} f(x; \Theta_0) T \, d\mu(x)$$
$$= 1, \qquad \text{if } i_1 = 1, i_2 = \cdots = i_k = 0,$$
$$= 0, \qquad \text{otherwise.}$$

Letting S_β stand for any one of the above defined S's except S_1 and applying the lemma of Section 2, the following inequality is obtained (subject to Assumption A) for the variance of the unbiased estimate T:

$$\text{var}_{\Theta_0} T \geqq \inf_{\substack{\theta_1^{(1)}, \theta_1^{(2)}, \cdots \\ \cdots\cdots\cdots \\ \theta_k^{(1)}, \theta_k^{(2)}, \cdots}} \frac{1}{\sigma^2_{S_1 \cdot S_{\beta'} S_{\beta''} \cdots S_{\beta^{(t)}}}}. \tag{3.2}$$

If the usual regularity conditions are assumed it is easily seen that this bound is at least as large as the ordinary Bhattacharyya bound.

For a biased estimate T having $E_\Theta(T) = g(\Theta)$, the following inequality is obtained (subject to Assumption A):

$$\text{var}_{\Theta_0} T \geqq \inf_{\substack{\theta_1^{(1)}, \theta_1^{(2)}, \cdots \\ \cdots\cdots\cdots \\ \theta_k^{(1)}, \theta_k^{(2)}, \cdots \\ l_{\beta'}, l_{\beta''}, \cdots}} \frac{\left[\Delta_{\theta_1^{(1)}} g(\Theta_0) - \sum l_{i_1 \cdots i_k} \Delta^{i_1}_{\theta_1^{(1)} \cdots \theta_1^{(i_1)}} \cdots \Delta^{i_k}_{\theta_k^{(1)} \cdots \theta_k^{(i_k)}} g(\Theta_0)\right]^2}{\text{var}_{\Theta_0} (S_1 - \sum l_\beta S_\beta)}. \tag{3.3}$$

4. Multistatistic case. For more than one statistic, say $(T_1, \cdots, T_m) = \mathbf{T}$, there is an immediate generalization of the inequality (3.3). It is obtained from the covariance relation that $[\sum_{yy} - \sum_{yx} \sum_{xx}^{-1} \sum_{xy}]$ is positive semi-definite, where $\sum_{yy}$, $\sum_{xx}$ and $\sum_{xy}$ are respectively the covariance matrices for a vector $\mathbf{y}$, for a vector $\mathbf{x}$, and between the vectors $\mathbf{x}$ and $\mathbf{y}$. Letting $\mathbf{y}$ be the statistic and $\mathbf{x}$ be a set of the S's defined by (3.1), then $\sum_{xy}$ becomes a matrix of differences of $E_\Theta(\mathbf{T})$ (as in the numerator on the right hand side of (3.3)).

5. Binomial distribution. For the unbiased estimation of the parameter p of the binomial distribution, the following lower bound for the variance at p is obtained using S and an interval h for differencing:

$$\sigma_T^2 \geqq \frac{h^2}{\left[1 + h^2 \dfrac{1}{p_0 q_0}\right]^n - 1}.$$

The greatest lower bound is obtained by letting $h \to 0$.

This greatest lower bound can also be obtained by using the Bhattacharyya bound (3.2) without applying a limiting operation:

$$\frac{p_0 q_0}{n} = \frac{1}{\sigma^2_{S_1 - S_2/2 + \cdots - (-1)^n S_n/n}},$$

where the S's are defined (3.1) as the divided differences corresponding to ordinary differences with interval h (h being chosen sufficiently small that all p's fall between 0 and 1).

REFERENCES

[1] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 581–586.
[2] E. L. Lehmann, "Notes on the Theory of Estimation," mimeographed notes.

ON THE ANALYSIS OF SAMPLES FROM k LISTS[1]

By Leo A. Goodman

The University of Chicago

1. Introduction and summary. Suppose we have k lists of names, no name appearing more than once in each list. We are interested in estimating the following parameters: (a) the number of names occurring in common in pairs, triples, $\cdots$, of lists; (b) the number of names occurring in $1, 2, \cdots, k$ lists. This note presents unbiased estimators for these parameters when a random sample is drawn from each list. It is also observed that the estimators presented are the only real-valued statistics which are unbiased estimators of the parameters, and hence must be the minimum variance unbiased estimators. This yields another example in which "insufficient" statistics have been used to obtain minimum variance unbiased estimators.

These unbiased estimators may at times give unreasonable estimates. In such cases, it is suggested that the statistics be modified so that the nearest reasonable estimate is used. Although this procedure introduces some bias, it usually reduces the mean square error.

This problem arises when we are interested in tracing the interrelations of agencies through the individual members. The problem also arises in the work of H. H. Fussler and J. M. Dawson of the University Library, University of Chicago, who are interested in comparing the acquisitions of various libraries. For special problems other sampling schemes may be more economical or more efficient than taking a sample from each list. Professor F. F. Stephan of Princeton University pointed out to the author that, in the special case of the "library problem," the Book Catalog and author cards used by many libraries provide a convenient means of drawing matched samples. (There is a brief discussion

[1] This work was prepared in connection with research supported by the Office of Naval Research.

of this kind of sampling problem on page 571 of [1].) A sampling scheme based on the last digit or two of the serial number of the cards could be used to search each library reference file for the same list of books. Special provision must be made for accessions made outside the sampling period and for books not covered by the Library of Congress cards. The analysis presented herein deals with the case in which (either for good, bad, or no reasons) a random sample has been drawn from each list.

The restriction that no name appear more than once in each list may be weakened to obtain somewhat more general results.

The problem discussed in this paper was brought to the author's attention by Professor W. Allen Wallis of the University of Chicago.

2. Results.

THEOREM 1. *Given k lists of names, let d_{12} names occur in common in lists 1 and 2, d_{13} names occur in common in lists 1 and 3, $\cdots$, $d_{[t]}$ names occur in common in lists $[t]$ (where $[t]$ is some subset containing at least two of the integers $1, 2, \cdots, k$), $\cdots$, $d_{123\cdots k}$ names occur in all lists. Suppose a random sample of $n_i = N_i/g_i$ names is drawn from list i, which contains N_i names, for $i = 1, 2, \cdots, k$. If e_{12} names occur in common in the samples from lists 1 and 2, e_{13} names occur in common in the samples from lists 1 and 3, $\cdots$, $e_{[t]}$ names occur in common in the samples from lists $[t]$, $\cdots$, $e_{12\cdots k}$ names occur in all samples, then an unbiased estimator of $d_{[t]}$ is*

$$d^1_{[t]} = \prod_i g_i e_{[t]},$$

where the product is taken over all values of i appearing in $[t]$.

The proof is based on the fact that $e_{[t]} = \sum \delta_{j[t]}$, where

$$\delta_{j[t]} = \begin{cases} 1, \text{ if name } j \text{ appears in all the samples from lists } [t] \\ 0, \text{ otherwise,} \end{cases}$$

and the summation is taken over all names.

THEOREM 2. *An unbiased estimator of the number of names occurring in ν lists is*

$$\sum_{i=0}^{k-\nu} (-1)^i C_\nu^{\nu+i}\, d^1(\nu + i),$$

where $d^1(\nu + i) = \sum d^1_{[t]}$ and the summation is taken over all $[t]$ containing $\nu + i$ integers. Also, an unbiased estimator of the number of names occurring in at least ν lists is

$$\sum_{i=0}^{k-\nu} (-1)^i C_{\nu-1}^{\nu+i-1}\, d^1(\nu + i).$$

The proof of these results follows from Theorem 1 and some combinatorics.

THEOREM 3. *Let F be a real-valued function of the parameters d_{12}, d_{13}, $\cdots$, $d_{[t]}$, $\cdots$, $d_{12\cdots k}$. Then there can be at most one real-valued function S of the sample results e_{12}, e_{13}, $\cdots$, $e_{[t]}$, $\cdots$, $e_{12\cdots k}$, such that $E\{S\} = F$, for all values of the parameters.*

PROOF. Let $2^k - k - 1 = M$. Suppose we order the M subsets $\{[t]\}$. To simplify notation we shall designate $d_{[t]}$ and $e_{[t]}$ by d_i and e_i, respectively, where $i = i([t])$ is the rank of the ordered subset $[t]$. The sample space consists of a subset $\{[e_1, e_2, \cdots, e_M]\}$ of M-dimensional Euclidean space. Let us order this subset by increasing values of e_M; for equal values of e_M, we order the vectors by increasing values of $e_{M-1}, \cdots$, for equal values of e_2, we order the vectors by increasing values of e_1. Hence, we may describe the sample space as a sequence $O_1 = [e_1(1), e_2(1), \cdots, e_M(1)]$, $O_2 = [e_1(2), e_2(2), \cdots, e_M(2)], \cdots$, where O_1 is the smallest ordered vector, O_2 is the next smallest, etc. To each sample point O_j let correspond the parameter point $P_j = [d_1(j), d_2(j), \cdots, d_M(j)]$, where $d_i(j) = e_i(j)$. Let $\Pr\{O_i ; P_j\}$ be the probability of obtaining sample point O_i when P_j is the true parameter point. Then it is easy to see that $\Pr\{O_i ; P_j\} = 0$ for $i > j$ and $\Pr\{O_i ; P_i\} > 0$. Hence, any unbiased estimate $S(O_i)$ of a function $F(P)$, defined on the parameter space P, must be such that

$$\sum_{i=1}^{j} S(O_i)\, Pr\{O_i ; P_j\} = F(P_j)$$

for $j = 1, 2, 3, \cdots$. This necessary condition insures the uniqueness of $S(O_i)$, since $S(O_i)$ must satisfy the recursion relation associated with the necessary condition.

In order to calculate the variance of these statistics, we again consider the estimators in terms of δ's. We then see that the variance of $d'[t]$ is

$$\sigma^2_{d_{[t]}} \sim d_{[t]} \prod_i g_i ,$$

where the product is taken over all values of i appearing in $[t]$, which permits the calculation of standard errors for the estimators. Similar results may be obtained for the other estimators presented.

By Theorem 3 we see that if one wishes to have unbiased estimators, then using the results of Theorems 1 and 2 is the best possible move. That is, the statistics described in those theorems are the only unbiased estimators of the parameters, and hence must be minimum variance unbiased estimators. The reader may have observed that $e_{[t]}$ is not a sufficient statistic for $d_{[t]}$. We see, therefore, that minimum variance unbiased estimators have been obtained using statistics which are not sufficient.

REFERENCE

[1] FREDERICK F. STEPHAN, "Practical problems of sampling procedure," *Am. Sociol. Rev.*, Vol., 1 (1936), pp. 569–580.

THE STANDARD ERROR OF GINI'S MEAN DIFFERENCE

By Z. A. Lomnicki

Polish University College, London

A general expression for the standard error of Gini's mean difference g was given in a paper under the same title by U. S. Nair [1]. See also [2], pp. 216–217.

The object of this note is to deduce in a more direct way a simpler formula for the variance of this statistic. The expression obtained is equivalent to that given by Nair except for an additional term overlooked in his final formula. The simplification is due to the fact that, for the evaluation of the expected values of g and g^2, it is not necessary to arrange the sample values in ascending order of magnitude as done by Nair.

Let n be the size of the sample, $f(x)$ the probability density function of the parent population, μ the mean and σ^2 the variance of x in the parent and let

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \qquad Z(x) = \int_{-\infty}^{x} tf(t)\,dt.$$

From the definition

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} | x_i - x_j | \tag{1}$$

(where the values x_i are not in order of magnitude but are numbered as they appear in the sample), we have

$$\begin{aligned} E(g) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(| x_i - x_j |) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} | x - y | f(x)f(y)\,dx\,dy = \Delta, \end{aligned} \tag{2}$$

where Δ is the mean difference (parameter) of the parent population. It is easy to check that Δ can also be written

$$\Delta = 2 \int_{-\infty}^{+\infty} \{xF(x) - Z(x)\} f(x)\,dx = 2 \int_{-\infty}^{\infty} xf(x)(2F(x) - 1)\,dx. \tag{3}$$

In order to find $E(g^2)$ let us write

$$\begin{aligned} g^2 = \frac{4}{n^2(n-1)^2} \{ \textstyle\sum (x_i - x_j)^2 &+ 2\textstyle\sum | x_i - x_j | \, | x_i - x_k | \\ &+ 2\textstyle\sum | x_i - x_j | \, | x_k - x_l | \}. \end{aligned} \tag{4}$$

The first sum should be read as the double sum extended to all pairs of different subscripts i, j, and has $n(n-1)/2$ terms; the second as a triple sum extended to all combinations of two pairs (i, j), (i, k) of different subscripts i, j, k and has $n(n-1)(n-2)/2$ terms; the third as a quadruple sum extended to all combinations of two pairs (i, j), (k, l) of different indices i, j, k, l and has $n(n - 1)(n - 2)(n - 3)/8$ terms. Thus

$$(5) \quad E(g^2) = \frac{1}{n(n-1)} \{2E(x_i - x_j)^2 + 4(n-2)E(|x_i - x_j| \, |x_i - x_k|) + (n-2)(n-3)E(|x_i - x_j| \, |x_k - x_l|)\}.$$

The first expected value is equal to $2\sigma^2$; the third to Δ^2. Denoting the second by J we have

$$(6) \quad \operatorname{var}(g) = E(g^2) - \Delta^2 = \frac{1}{n(n-1)} (4\sigma^2 + 4(n-2)J - 2(2n-3)\Delta^2),$$

where

$$(7) \quad J = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| \, |x - z| \, f(x)f(y)f(z) \, dx \, dy \, dz.$$

This can be written as

$$(8) \quad \begin{aligned} J = \int_{-\infty}^{\infty} f(x) \Big\{ &\int_{-\infty}^{x} \int_{-\infty}^{x} (x - y)(x - z)f(y)f(z) \, dy \, dz \\ &+ \int_{-\infty}^{x} \int_{x}^{\infty} (x - y)(z - x)f(y)f(z) \, dy \, dz \\ &+ \int_{x}^{\infty} \int_{-\infty}^{x} (y - x)(x - z)f(y)f(z) \, dy \, dz \\ &+ \int_{x}^{\infty} \int_{x}^{\infty} (y - x)(z - x)f(y)f(z) \, dy \, dz \Big\} dx. \end{aligned}$$

Putting

$$(9) \quad G(x) = \int_{-\infty}^{x} (x - y)f(y) \, dy = xF(x) - Z(x),$$

$$(10) \quad H(x) = \int_{x}^{\infty} (y - x)f(y) \, dy = G(x) + \mu - x,$$

we obtain

$$(11) \quad \begin{aligned} J &= \int_{-\infty}^{\infty} [G^2(x) + 2G(x)H(x) + H^2(x)]f(x) \, dx \\ &= \int_{-\infty}^{\infty} [G(x) - H(x)]^2 f(x) \, dx + 4 \int_{-\infty}^{\infty} G(x)H(x)f(x) \, dx, \end{aligned}$$

and finally

$$(12) \quad \operatorname{var}(g) = \frac{1}{n(n-1)} \{4(n-1)\sigma^2 + 16(n-2)I - 2(2n-3)\Delta^2\},$$

where

$$
(13) \qquad \begin{aligned} I &= \int_{-\infty}^{\infty} G(x)H(x)f(x)\,dx \\ &= \int_{-\infty}^{\infty} \{[xF(x) - Z(x)]^2 + (\mu - x)[xF(x) - Z(x)]\}f(x)\,dx. \end{aligned}
$$

This integral can also be written as

$$
(14) \qquad I = \int_{-\infty}^{\infty}\int_{-\infty}^{x}\int_{x}^{\infty} (x - y)(z - x)f(x)f(y)f(z)\,dx\,dy\,dz,
$$

and, according to the distribution involved, formula (13) or (14) may be more convenient in the evaluation of var (g).

Comparing (12) with the formulae given by Nair it is easy to show that an additional term $(n - 3)\mu^2$ has been omitted in his final formula for I_1. However, the values of var (g) for normal, exponential and rectangular distributions given in [1] are correct and agree with those obtained from formula (12) above.

REFERENCES

[1] U. S. Nair, "The standard error of Gini's mean difference," *Biometrika*, Vol. 28 (1936), pp. 428–436.

[2] M. G. Kendall, *Advanced Theory of Statistics*, Vol. I, Charles Griffin and Co., London 1943.

CORRECTION TO "A NOTE ON THE POWER OF A NONPARAMETRIC TEST"

By F. J. Massey, Jr.

University of Oregon

In the paper mentioned in the title (*Annals of Math. Stat.*, Vol. 21 (1950), pp. 440–443) the proof of the biasedness of a test based on the maximum deviation between sample and population cumulatives is incorrect. A proof is given below. Also, on page 442, line 2, "greater" should be replaced by "less". The notation refers to Fig. 1 of the original article.

Above point b (note $F_1(b) = F_0(b)$), there will be certain possible heights for $S_n(x)$ to attain and still remain in the band. Call these heights $b_1 = 1/n$, $b_2, b_3, \cdots, b_k = k/n$, where $k/n < 2d/\sqrt{n}$. Locate the point $x = c$ $(c < b)$ close enough to $x = b$ so that $F_0(c) + d/\sqrt{n} > b_k$. Then consider

$$
\text{(i)} \qquad P_0 = P\{S_n(x) \text{ remain in band} \mid F(x) = F_0(x)\},
$$

$$
\text{(ii)} \qquad P_1 = P\{S_n(x) \text{ remain in band} \mid F(x) = F_1(x)\}.
$$

Now $P_j = \sum_{i=1}^{k} P\{S_n(x)$ passes through b_i and remains in band $\mid F_j(x)\} = \sum_{i=1}^{k} P\{S_n(x)$ goes through $b_i \mid F_j(x)\} \cdot P\{S_n(x)$ stays in band for $x < b \mid F_j(x)$,

$S_n(x)$ goes through $b_i\} \cdot P\{S_n(x)$ stays in band for $x > b \mid F_j(x)$, $S_n(x)$ goes through b_i and is in band for $x < b\}$. However the first and third of the factors is the same for $j = 0, 1$, and the second is unity for $j = 1$, and therefore $P_0 \leqq P_1$. If $\lambda/\sqrt{N} > 1/N$ (which is necessary if the test is not always going to reject) then at least for height b_k,

$$P\{S_n(x) \text{ inside the band for } x < b \mid S_n(b) = b_k,\ F_0(x)\} < 1.$$

Thus the test is biased.

I would like to thank Professor D. A. Darling for pointing out the error.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the East Lansing meeting of the Institute, September 2-5, 1952)

1. An Extension of Massey's Distribution of the Maximum Deviation between Two Sample Cumulative Step Functions. (Preliminary Report.) CHIA KUEI TSAO, Wayne University.

Let $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_m$ be the ordered observations of two random samples from populations having cumulative distribution functions $F(x)$ and $G(x)$ respectively. Let $S_n(x) = k/n$ where k is the number of observations of X which are less than or equal to x and $S'_m(x) = j/m$ where j is the number of observations of Y which are less than or equal to x. The statistics $d_r = \max \mid S_n(x) - S'_m(x) \mid$ (max over $x < x_r$) and $d'_r = \max \mid S_n(x) - S'(x) \mid$ (max over $x < \max(x_r, y_r)$) can be used to test the hypothesis $F(x) = G(x)$. For example, using d_r we would reject the hypothesis if the observed value of d_r is significantly large. In this paper, the methods of obtaining the distributions of d_r and d'_r (for small size samples) are similar to that in Massey's paper, and several short tables for equal size samples are included. (Work supported by the Office of Naval Research.)

2. Polynomial Correlation Coefficients. W. D. BATEN AND J. S. FRAME, Michigan State College.

In this paper is developed a formula for the correlation coefficient pertaining to predicting polynomials. It is shown, when the independent variates are approximately normally distributed, that the square of this correlation coefficient can be expressed as a finite sum involving the squares of the averages of the derivatives of the estimating polynomial, namely, $r^2 = \Sigma \overline{y^{(k)}}^2/k!$, where y represents the predicting polynomial. The proof is based upon manipulations of Bernoulli numbers.

3. Truncated Poisson Distributions. PAUL R. RIDER, Wright-Patterson Air Force Base and Washington University.

This paper gives a method for estimating the parameter of truncated Poisson distributions for which some of the data are missing, particularly those which are truncated at the lower end. Application to a number of actual distributions is discussed.

4. Frequency Distributions for Functions of Rectangularly Distributed Random Variables. STUART T. HADDEN, Socony-Vacuum Laboratories, Paulsboro, New Jersey.

The theory of rectangularly distributed random variables is presented. It is shown how such random variables can occur in a certain class of controlled experiments arising in the fields of physics, chemistry, and engineering. On the basis of rectangularly distributed random variables, arising in observations or process variables, frequency distributions are developed for quantities which are functions of such variables. The principal method used in deriving the frequency distributions is operationally by means of the Laplace transform. Example applications illustrate how such frequency distributions can be applied in the analysis of experimental variance.

5. On Truncated Rules of Action. (Preliminary Report.) BENJAMIN EPSTEIN, Wayne University.

A rule of action of theoretical and practical interest in life testing can be described as follows: (a) Non-replacement. Start the life test with n items drawn from a population. Let an integer r_0 and truncation time T_0 be preassigned. By the nature of the experiment failures will occur in order. Let $X_{r_0,n}$ be the time when the r_0th ordered failure occurs. If $X_{r_0,n} < T_0$, stop the experiment at $X_{r_0,n}$ and take action I. If $X_{r_0,n} > T_0$, stop the experiment at T_0 and take action II. (b) Replacement. Same as non-replacement except that a failed item is replaced at once by a new item. The properties of this kind of rule are investigated in detail when the underlying pdf is of the form $(1/\theta)e^{-x/\theta}$, $x > 0$, a distribution of some interest in life testing. The distributions of r, the number of items destroyed before taking an action, and T, the length of the experiment, are obtained. In particular $L(\theta)$, the probability of taking action I (say), $E_\theta(r)$, and $E_\theta(T)$ are obtained. Some tables based on this theory are obtained. (Work supported by the Office of Naval Research.)

6. The Distribution of the Difference of Two Independent Chi-Squares. JAMES PACHARES, University of North Carolina.

As a special case of the problem of the distributions of quadratic forms being investigated by the author, let $T_n = X_n - Y_n$, where X_n and Y_n are independently and identically distributed with probability density function (pdf) $[\Gamma(n/2)]^{-1}e^{-u}u^{(n-2)/2}$, $u > 0$. If $f_n(t)$ denotes the pdf of T_n, then the following recurrence equation holds: $f_{n+4}(t) = \{(n+1)/(n+2)\} f_{n+2}(t) + \{1/[n(n+2)]\}t^2 f_n(t)$, $n = 1, 2, \cdots$. The exact distribution of T_n is derived. If $K_n(t)$ is the modified Bessel function of the second kind of order n, then $f_n(t) = \pi^{-\frac{1}{2}}[\Gamma(n/2)]^{-1} (|t/2|)^{(n-1)/2}K_{(n-1)/2}(|t|)$, $n = 1, 2, \cdots$. Recurrence relations between the cumulative distribution functions (cdf's) of T_n are established so that any cdf for odd n depends on $F_1(x)$, while any cdf for even n depends on $F_2(x)$, where $F_n(x) = Pr[|T_n| \leq x]$. A method is given for evaluating $F_1(x)$ by a series, with bounds on the error committed by stopping with a given term. Upper and lower bounds for $F_n(x)$ are given. (Work sponsored by the Office of Naval Research.)

7. Partially Balanced Designs with Two Plots per Block. R. C. BOSE, University of North Carolina, AND K. R. NAIR, University of North Carolina and Forest Research Institute, Dehradun, India.

In many experimental situations, the block size is compulsorily restricted to two, as in comparing treatments given to two halves of a leaf. Partially balanced designs requiring only a small number of replications and with m accuracies $m \leq 4$ have been worked out. It has been noticed that the association schemes of any known partially balanced incomplete block design with block size greater than 2 will lead to a design of the same type with block size 2, but a larger number of replications.

8. Minimax Sampling and Estimation in Finite Populations. Om Prakash Aggarwal, Stanford University.

Stratified and cluster sampling from a finite population is considered from Bayes and minimax point of view. The loss in estimating the mean is taken to be the cost of observations plus the squared error in the estimate of the mean. For stratified sampling with linear cost function, for instance, it is shown that the minimax sampling plan chooses $n_i = \left\{\sqrt{(N_i^2\sigma_i^2/c_i) + \frac{1}{4}}\right\}$ individuals at random from the ith stratum and uses the usual estimate $f = \Sigma_{i=1}^k N_i\overline{X}_i$ for estimating $\Sigma_{i=1}^k N_i\mu_i$, where k is the number of strata, and in the ith stratum, N_i denotes the total number of individuals, μ_i, σ_i^2, the mean and variance, c_i the cost of sampling per individual, $\overline{X}_i$ the sample mean, and $\{q\}$ the integer nearest to q.

9. Some Two Sample Tests on the Exponential Distribution. (Preliminary Report.) Benjamin Epstein and Chia Kuei Tsao, Wayne University.

Let S_{1n_1} and S_{2n_2} be two random samples such that S_{in_i} is a sample of size n_i from a population having pdf $(1/\theta_i)$ exp $[-(x - A_i/\theta)]$ $(i = 1, 2)$. Let S_{ir_i} be the set of the first $r_i(r_i \leqq n_i)$ smallest observations in S_{in_i}. On the basis of S_{1r_1} and S_{2r_2}, various likelihood ratio tests about the parameters involved can be obtained. The likelihood ratio tests about the hypothesis $\theta_1 = \theta_2$ assuming either A_1 and A_2 known or unknown are reducible to the well-known F-test. The test criterion for the hypothesis $A_1 = A_2$, assuming θ_1 and θ_2 known, may be reduced to a random variable having an exponential distribution. The tests of the hypothesis that $A_1 = A_2$ assuming θ_1 and θ_2 unknown, are also reduced to F-tests. Finally the test of the hypothesis $A_1 = A_2$ and $\theta_1 = \theta_2$ is obtained for the special case $r_1 = r_2$. (Work supported by the Office of Naval Research.)

10. Efficiency of Estimators of the Mean of an Exponential Distribution Based Only on the rth Smallest Observation in an Ordered Sample. Benjamin Epstein, Wayne University.

Let us assume that the lives of certain items are describable by a positive random variable X, whose pdf is $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $x > 0$. A sample of size n is drawn, and we suppose that the observations become available in order. Let the experiment be terminated at $x_{r,n}$, the time of failure of the rth item. We raise the question: How much information is lost if we base our estimate of the unknown parameter θ only on $x_{r,n}$ instead of basing it on all the first r failure-times, $x_{i,n}$, $i = 1, 2, \cdots, r$? As reported recently the m. l. estimate based on the $x_{i,n}$ is given by $\hat{\theta}_{r,n} = U/r$ where $U = \Sigma_{i=1}^r x_{i,n} + (n - r)x_{r,n}$. This estimate is "best" in the sense that it is unbiased, minimum variance, efficient, and sufficient. It is shown that unbiased estimates of θ based on $x_{r,n}$ alone have high efficiencies ($\geqq .9$) relative to $\hat{\theta}_{r,n}$ for values of $r \leqq 2n/3$. For example, for $r = n/2$, $n =$ even integer, the efficiency $\geqq 2(\log 2)^2 = .9608$. Tables giving the unbiasing constants $\beta_{r,n}$ such that $E(\beta_{r,n}x_{r,n}) = \theta$, $\text{Var}(\beta_{r,n}X_{r,n})$, and the efficiencies $\text{Var}(\hat{\theta}_{r,n})/\text{Var}(\beta_{r,n}X_{r,n})$ have been obtained for $n = 1(1)20(5)30(10)100$ and $r = 1(1)n$. (Work supported by the Office of Naval Research.)

11. On the Theory of Systematic Sampling. III. William G. Madow, University of Illinois.

It is shown that if the elements of the population are constants and the population is monotone then centered systematic sampling is more efficient than random start systematic sampling; and that if the elements of the population are random variables and the correlogram is monotone decreasing then centered systematic sampling is more efficient than random start systematic sampling while if the correlogram is monotone increasing the contrary is true.

12. The Power of Some Service Tests. Leo A. Goodman, University of Chicago.

George W. Brown and Merrill M. Flood have presented in an interesting paper ("Tumbler Mortality", *Jour. Am. Stat. Assn.*, Vol. 42 (1947), pp. 567–574) the results of an analysis of a service test that was used to determine which of two types of glass tumblers had a longer mean length of life when used in a particular cafeteria. At the end of each week, each broken tumbler was recorded and replaced by a new one of the same type. Another kind of service test is based on the procedure of replacing the tumblers in equal numbers; i.e., as many of type 1 as of type 2, even though they broke in unequal numbers. Still another kind of service test is based on the procedure of replacing each broken tumbler by a new one of the other type. The preceding two procedures suggested, may be performed using either weekly records, or only the final count (the latter is less powerful, but less work). The exact power of these service tests is computed under the assumption of constant risk. The asymptotic power is computed in the more general case (non-constant risk). The several service tests are compared. This information may be used by the experimenter to decide which one of these tests to perform, and when to conclude the test.

13. A Minimal Essentially Complete Class of Tests of a Simple Hypothesis Specifying the Mean of a Unit Rectangular Distribution. Allan Birnbaum, Columbia University.

For the problem of testing a simple hypothesis on the mean of a unit rectangular distribution, on the basis of n ($n \geqq 2$) observations, explicit characterizations of the minimal complete class and a minimal essentially complete class of tests are given. Examples of tests which are best against various classes of alternatives are given; it is shown that the test with highest power against alternatives far from the null hypothesis has minimum power against alternatives close to the null hypothesis.

14. Application of Random Walk Theory to a General Class of Sequential Decision Problems. (Preliminary Report.) G. E. Albert, University of Tennessee.

One of r decisions d_i , $i = 1, 2, \cdots, r$, is to be made concerning the conditional cdf $F(y \mid x)$ of y given x, x and y in a Euclidean space R, by the following sequential experiment. Assign $r + 1$ nonnegative functions $p_i(x)$, $i = 0, 1, 2, \cdots, r$, on R with $\sum_{i=0}^{r} p_i(x) = 1$. Perform a random walk beginning at an arbitrary point x_0, with successive points x_j drawn from $F(x_{j+1} \mid x_j)$, $j = 0, 1, 2, \cdots$, and terminating as soon as one of d_i, $i = 1, 2, \cdots, r$, has been decided under the following rule: let d_0 denote the decision to continue experimentation after any step x_j of the walk; at each step x_j, $j = 0, 1, 2, \cdots$, one of the decisions d_i, $i = 0, 1, 2, \cdots, r$, is made with respective probabilities $p_i(x_j)$. Let $P_i(x)$ denote the probability of making the decision d_i, $i = 1, 2, \cdots, r$, as a result of a walk starting at x. It is shown that under certain mild restrictions $P_i(x) = p_i(x) + p_0(x) \int_R P_i(y)\, dF(y \mid x)$. Also, the moment generating function and the moments of the duration of the experiment satisfy integral equations of a similar type; see Wasow, "On the duration of random walks." *Annals of Math. Stat.*, Vol. 22 (1951), pp. 199–216, for a special case. Some methods of approximating the solutions of these integral equations are established. Application of the theory is illustrated by a discussion of the sequential probability ratio test of hypotheses

θ_1, θ_2 on the parameter θ of a general class of cdf $G(x; \theta)$ which possesses a sufficient statistic for the parameter.

15. Nonparametric Comparisons of Populations when Data Are Collected in Homogeneous Groups. Frank J. Massey, Jr., University of Oregon.

The method of comparing two populations when data are paired has been fairly widely studied; for example, the sign test or t-test on differences. This paper presents similar techniques for analyzing data which have been collected in groups of size larger than one from each of several populations. Comparisons of power curves are made for certain normal alternatives (Work sponsored by the Office of Naval Research.)

16. On the Reduced Moment Problem. Salem H. Khamis, Statistical Office, United Nations.

Let $\Phi(x)$ and $\Psi(x)$, $a \leqq x \leqq b$, be any two distinct cumulative distribution functions which are continuous and differentiable solutions of the reduced moment problem $\mu_r = \int_a^b x^r d\alpha(x)$, $r = 0, 1, 2, \cdots, 2n$. A proof is given of the inequality (1) $| \Phi(x) - \Psi(x) | \leqq K\rho_n(x)$, where $\rho_n(x) = -| \mu_{i+j} |/D_n(x)$, $i, j = 0, 1, 2, \cdots, n$, $D_n(x)$ is the determinant obtained by bordering the determinant $| \mu_{i+j} |$ by the prefixed row $(0\ 1\ x\ x^2 \cdots x^n)$ and the corresponding column, and where $0 < K = AB/(A + B - AB) \leqq \text{Min}(A, B) \leqq 1$ with $0 < A = 1 + \text{l.u.b.}_{a \leqq x \leqq b}(-\Phi'(x)/\Psi'(x)) \leqq 1$ and $0 < B = 1 + \text{l.u.b.}_{a \leqq x \leqq b}(-\Psi'(x)/\Phi'(x)) \leqq 1$. Inequality (1) is an improvement of an earlier result by the same author (*Proceedings of the International Congress of Mathematicians*, 1950, Vol. I, p. 569) which is in turn an improvement upon the corresponding Tchebycheff inequality, i.e., without the constant K (Shohat and Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943, p. 72). By a special differencing method it is shown that the magnitude of the determinant in the numerator of $\rho_n(x)$ is independent of the origin of the moments, and that the determinant in the denominator is expressible in terms of the moments about the origin x. A method is also given for constructing an infinite number of cumulative distribution functions defined over a finite interval and possessing equal moments up to any given order, making use of the properties of orthogonal polynomials. Inequality (1) is then applied to the special class of such cumulative distribution functions associated with the Legendre polynomials.

17. Canonical Partial Correlations. S. N. Roy and J. Whittlesey, University of North Carolina.

Canonical partial correlations between a set of p and a set of q variates, after elimination of a third set of r variates, is obtained by considering the canonical correlations between a set of $(p + r)$ and a set of $(q + r)$ variates having r variates in common. Suppose S is a $(p + q + r) \times (p + q + r)$ p.d. covariance matrix partitioned into submatrices such that the first row is $S_{11}(:p \times p)$ $S_{12}(:p \times q)$ $S_{13}(:p \times r)$, the second row is $S'_{12}(:q \times p)$ $S_{22}(:q \times q)$ $S_{23}(:q \times r)$, and the third row is $S'_{13}(:r \times p)$ $S'_{23}(:r \times q)$ $S_{33}(:r \times r)$. Then the canonical partial correlations between the p set and the q set are given by the p nonnegative roots (all lying between 0 and 1) of the equation in θ:

$$| \theta(S_{11} - S_{13}S_{33}^{-1}S_{13}) - (S_{12} - S_{13}S_{33}^{-1}S'_{23})(S_{22} - S_{23}S_{33}^{-1}S'_{23})^{-1}(S'_{12} - S_{23}S_{33}^{-1}S'_{13}) | = 0.$$

Putting (i) $r = 0$, (ii) $p = 1$, (iii) $p = 1, q = 1$, and (iv) $p = 1, r = 0$, we have respectively (i) canonical correlations, (ii) multiple partial correlation, (iii) partial correlation, and (iv) multiple correlation.

18. A Useful Transformation in the Case of Canonical Partial Correlations. S. N. Roy, University of North Carolina.

If the distribution problem in Abstract 31 were to be tackled ab initio, that is, without assuming the distributions of canonical correlations, the following transformation would be very helpful: $X_1(:p \times n) = U_1(:p \times p) \times (D_{\sqrt{1-\theta}} D_{\sqrt{\theta}}) \times$ (a $2p \times n$ matrix whose first row is $L_1(:p \times n)$ and the second row is $L_3(:p \times n)) + V_1(:p \times r)L_4(:r \times n)$. Also $X_2(:q \times n) =$ (a $q \times q$ matrix partitioned into 4 submatrices such that the first row is $U_{21}(:\overline{q-p} \times p)\tilde{U}_{22}(\overline{q-p} \times \overline{q-p})$ and the second row is $U_{23}(:p \times p)U_{24}(:p \times \overline{q-p})) \times$ (a $q \times n$ matrix whose first row is $L_2(:p \times n)$ and the second row is $L_3(:\overline{q-p} \times n)) + V_2(:q \times r)L_4(:r \times n)$, and lastly $X_3(:r \times n) = \tilde{U}_3(:r \times r)L_4(:r \times n)$ where the $(p + q + r) \times n$ matrix X, which is the reduced matrix of observations is supposed to be partitioned into $X_1(:p \times n)$, $X_2(:q \times n)$ and $X_3(:r \times n)$ placed one below the other, D_a stands for a diagonal matrix with elements $(a_1, \cdots, a_p)$, θ is given by the equation in the above abstract, $\tilde{M}$ stands for any triangular matrix with upper right hand corner zero, and $L'(:n \times \overline{p + p + q - p + r}) = (L_1'L_2'L_3'L_4')$ is subject to $LL' = I(p + q + r)$. This transformation for an X of rank $p + q + r$ can be shown to exist and could also be made one to one if (i) the θ's are distinct, and, say, (ii) the first row of U_1, the diagonal elements of $\tilde{U}_{22}$ and of $\tilde{U}_3$ are all taken to be positive. This will of course happen almost everywhere (in the sample space). Erasing X_3, $\tilde{U}_3$, L_4, V_1 and V_2 we have the case of canonical correlations.

19. Uniform Convergence of Distribution Functions. Emanuel Parzen, University of California, Berkeley.

We determine conditions under which uniform convergence in a parameter θ of sequences of characteristic functions implies uniform convergence in θ of the corresponding sequences of distribution functions, which may be univariate or multivariate. We then derive a uniform central limit theorem and a uniform weak law of large numbers for sequences of independent random variables whose distribution depends on θ. These results may be applied to obtain conditions for the uniform consistency and uniform asymptotic normality of maximum likelihood estimates to be compared with those given by A. Wald ("Asymptotically most powerful tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 12 (1941), p. 2).

20. Statistical Aspects of a Linear Programming Problem. D. F. Votaw, Jr., Yale University.

The Hitchcock-Koopmans transportation problem is to determine a most economical program of transporting a homogeneous product (e.g., oil) from origins to destinations. The amounts of the product at the origins and required at the destinations are given together with the cost of transporting a unit amount from any origin to any destination. This paper is concerned with the analogous problem arising when the costs are unknown parameters in a distribution from which a sample is available. An application of the analysis of variance is pointed out, and some results of synthetic sampling are presented. (Research sponsored by the Office of Naval Research.)

21. Maximum Likelihood Estimators and A Posteriori Distributions. J. Wolfowitz, Cornell University.

Let $f(x, \theta)$ be the frequency function at x of each of the independent chance variables $x_1, \cdots, x_n$, whose distribution depends upon the parameter θ. Let $g(\theta')$ be the a priori density function of θ at θ', and let $h(\theta' \mid x_1, \cdots, x_n)$ be the a posteriori density function of θ at θ', given $x_1, \cdots, x_n$. Under suitable regularity conditions on f and g, h is asymptotically normal, with mean $\hat{\theta}$ and variance $[nc(\hat{\theta})]^{-1}$, where $\hat{\theta}$ is the maximum likelihood estimate of θ from $x_1, \cdots, x_n$ and $c(\theta) = \int ((\partial \log f(x, \theta))/\partial\theta)^2 f(x, \theta)\, dx$. Thus the influence of g disappears in the limit. The present result includes that of v. Mises (*Math. Zeit.*, Vol. 4 (1919)) for the binomial case, and those of Kolmogoroff (*Izvyestya Akad. Nauk SSSR, Ser. Mat.*, Vol. 6 (1942)) for the normal case.

22. **Estimates and Asymptotic Distributions of Certain Statistics in Information Theory. (Preliminary Report.)** JOHN P. HOYT, U. S. Naval Academy.

In "On information and sufficiency" (S. Kullback and R. A. Leibler, *Annals of Math. Stat.*, Vol. 22 (1951), pp. 79–86), the concepts of "information" (designated hereafter as "i") and "mean information per observation" (designated hereafter as "I") for discrimination between two hypotheses were defined and various properties of "I" were proved for the abstract case. In "An application of information theory to multivariate analysis" (S. Kullback, *Annals of Math. Stat.*, Vol. 23 (1952), pp. 88–102), certain applications of information theory were made to multivariate analysis but problems of estimation and distribution were not considered. In the present paper, the characteristic function of the distribution of "i" in a sample of n from a normal multivariate population is found and from this is derived the expected value and variance of "i". A sample estimate of n "I" is then considered assuming equality of means in the two populations and a known value of one of the variance-covariance matrices occurring in "I". Using unbiased estimates of the parameters occurring in the other variance-covariance matrix, the characteristic function of the distribution of the estimate is found and is then used to show that the estimate's asymptotic distribution is given by the chi-square distribution with $k(k + 1)/2$ degrees of freedom.

23. **On Testing One Simple Hypothesis Against Another.** LIONEL WEISS, University of Virginia.

Given a sequence $(X_1, X_2, \cdots)$ of independently and identically distributed chance variables, H_0 is the hypothesis that the probability density function of each chance variable is $f_0(x)$, H_1 is the hypothesis that this function is $f_1(x)$. A "generalized sequential probability ratio test" is defined as the usual Wald sequential probability ratio test, except that constant limits are not necessarily used; in other words, after the i *th* observation is taken, accept H_0 if the probability ratio is not greater than B_i, accept H_1 if the ratio is not less than A_i, otherwise take another observation, where $0 \leqq B_i \leqq A_i$. Given any test T of H_0 against H_1, not using randomization, and such that the probability that T will terminate is 1 when either H_0 or H_1 is true, then under mild restrictions on $f_0(x)$ and $f_1(x)$ the following theorem holds: There exists a sequence $(G_1, G_2, \cdots)$ of generalized sequential probability ratio tests such that Pr(sample size, when using G_j and H_i is true, is no greater than n) $\geqq Pr$(sample size, when using T and H_i is true, is no greater than n) for all n, all integers j, and $i = 0, 1$; and also, as j approaches infinity, lim $Pr(H_i$ will be accepted when it is true and G_j is used) exists and is not less than $Pr(H_i$ will be accepted when it is true and T is used), for $i = 0, 1$. If T is a truncated test, a stronger theorem holds; there exists a generalized sequential probability ratio test, also truncated, enjoying the above advantages over T.

24. **Extreme Value Theory for m-Dependent Stationary Sequences of Continuous Random Variables.** GEOF. WATSON, University of Melbourne.

The distributions, and their limits as $N \to \infty$, of the order statistics of N successive observations in a sequence of independent continuous random variables with a common distribution function, are well known. The present paper considers the same problem for sequences governed by stationary m-dependent probability laws. A stationary sequence is called m-dependent if $P(x_0 \leqq k_0 \mid x_{-1} \leqq k_{-1}, \cdots) = P(x_0 \leqq k_0 \mid x_{-1} \leqq k_{-1}, \cdots, x_{-m} \leqq k_{-m})$. These distributions are found in the general case here and it is shown that, as $N \to \infty$, their limiting forms are the same as the distributions obtained in the case of independence provided max $\{P(x_i > k, x_j > k)/P(x > k)\} \to 0(k \to \beta, k \leqq \beta)$ and max $\{P(x_i \leqq k, x_j \leqq k)/P(x \leqq k) \to 0(k \to \alpha, k \geqq \alpha)$, where the maximum is taken for $i, j = 1, \cdots, m + 1, i \neq j$, and where (α, β) is the range of the random variables $x_t(t = \cdots, -1, 0, 1, \cdots)$. Either or both of α and β may be infinite. These latter conditions are shown to be satisfied in all stationary normal processes. Thus the results of this paper give the limiting distributions of the order statistics in a sample of successive observations from any normal stationary autoregressive process.

25. Sequential Tests and Estimates for Comparing Poisson Populations. Allan Birnbaum, Columbia University.

The problem of testing a hypothesis on $\gamma = \lambda_2/\lambda_1$ is considered, where λ_1, λ_2 are the means of two Poisson populations. It is shown that no nonsequential test of $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma = \gamma_1$ can have size uniformly $\leqq \alpha$ and power uniformly $\geqq 1 - \beta(1 - \beta > \alpha)$; a simple sequential test (not of the Wald type) is given which has these requirements of size and power against one- or two-sided alternatives. The generalization to the problem of classifying γ into one of k intervals is indicated. Comparisons with the Wald sequential tests of $H_0 : \gamma = \gamma_0$ against one-sided alternatives and of $H_0 : \gamma = 1$ against two-sided alternatives are made. The latter one-sided tests are constructed by use of a simple sufficient condition for the existence of a sequential probability ratio test of a composite hypothesis $H_0 : \theta \varepsilon \omega_0$, against a composite alternative $H_1 : \theta \varepsilon \omega_1$, of size approximately α for all $\theta \varepsilon \omega_0$ and power approximately $1 - \beta$ for $\theta \varepsilon \omega_1$. Application of this condition to problems of comparing two populations with Koopman-form distributions also gives tests which include those given by Girshick ("Contributions to the Theory of Sequential Analysis. I," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 123–143), and some new tests for comparing variances of two normal populations. Tests of equality of ratios of means of two pairs of Poisson populations are given.

26. Sequential Decision Problems in the Stationary Case. J. Kiefer, Cornell University.

Results of Wald and Wolfowitz ("Bayes' solutions of sequential decision problems," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 82–99; also, Chap. 4 of Wald's *Statistical Decision Functions*, John Wiley and Sons, 1950) are generalized to the case where the chance variables are no longer assumed independent, but instead form a stationary process. Questions of measurability and existence, recurrence formulas, characterizations of Bayes' solutions, etc., are simplified by first considering only nonrandomized decision functions and by then using results of the same authors ("Two methods of randomization in statistics and the theory of games," *Ann. Math.*, Vol. 53 (1951), pp. 581–586) to extend the conclusions to randomized procedures. The essential difference between the independent case and, e.g., the stationary Markoff case, is that in the latter a Bayes' solution may depend at each stage on the last observation as well as on the a posteriori distribution. For example, a Bayes' solution for testing between two simple hypotheses in the Markoff case is characterized by two functions $B(x) \leqq A(x)$ (which under slight restrictions are continuous) which are used after m observations the last of which is x_m by comparing the probability ratio to $B(x_m)$ and $A(x_m)$. Unlike the independent case, the $B(x)$ and $A(x)$ cannot in general be replaced by constants independent of x; nor does every pair $B(x)$, $A(x)$ constitute a Bayes' solution relative to some weight function, cost, and a priori distribution (as does every pair B, A in the independent case); nor need a Bayes' solution possess the optimum property of the independent case.

27. Random Functions Satisfying Certain Linear Relations. II. Sudhish G. Ghurye, University of North Carolina.

The particular case $k = 1$ of the problem mentioned in Part 1 is considered here in detail. Let $X(t)$ be a p-dimensional, real-valued random function, defined and continuous in probability for all t in an interval $[t_0, T]$. Further, let there exist a real-valued, $p \times p$ matric function $A(h)$, defined and continuous for $h > 0$, such that if we write $Y(k; h) = X(t_0 + kh) - A(h)X(t_0 + [k - 1]h)$, then for any positive h and any integer n ($nh \leqq T - t_0$), $X(t_0)$, $Y(1; h)$, $\cdots$, $Y(n; h)$ are mutually independent. Then it is shown that $A(h)$ can be written in the form e^{Bh}, where B is a constant matrix, and that $X^*(t) = e^{-Bt}X(t)$ is a random function with independent increments (r.f.i.i.). It is also shown that if $Z(t)$ is any p-dimensional r.f.i.i. and $A(t)$ is a $p \times p$ matric function, continuous and of bounded variation, then the integral $\int_{t_0}^{t} A(v)\, dZ(v)$ exists as the unique limit-in-distribution of the sequence of approximating sums. From this, a one-to-one correspondence (in distribution) between the random functions $X(t)$ mentioned above and the random functions $\int_{t_0}^{t} e^{(t-v)B}\, dZ(v)$ is established.

28. Optimal Designs for Estimating Parameters. (Preliminary Report.) Herman Chernoff, Stanford University.

The following is a generalization of a result of Elfving (see "Optimum allocation in linear regression theory," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 255–262). It is desired to estimate parameters $\theta_1, \theta_2, \cdots, \theta_s$. There is available a set of experiments which may be performed. The probability distribution of the data obtained from any of these experiments may depend not only on $\theta_1, \theta_2, \cdots, \theta_s$ but also on the nuisance parameters θ_{s+1}, $\theta_{s+2}, \cdots, \theta_k$. One is permitted to select a design consisting of n of these experiments to be performed independently. The repetition of experiments is permitted in the design. Then it can be shown that under mild conditions and for large n locally optimal designs may be approximated by selecting a set of $r \leqq k + (k - 1) + \cdots + (k - s + 1)$ of the experiments available and by repeating each of these r experiments in certain specified proportions. The criterion of optimality used is a natural one involving the information matrices of the experiments.

29. The Distribution of the nth Variate in Certain Chains of Serially Dependent Populations. L. V. Toralballa, Marquette University.

The following is a representative of the problems considered: Let $P_1, P_2, \cdots, P_n$ be a sequence of normally distributed populations, the first having a mean m_1 and variance σ_1^2, each population after the first having a mean $m_i = ax_{i-1} + b$, where x_{i-1} is a random value of the variate in P_{i-1} and a variance σ_i^2. One seeks the absolute distribution of the variate in P_n. In this particular case it is found that the absolute distribution of the variate in P_n is normal, with a mean $b\Sigma_0^{n-2} a^i + a^{n-1}m_1$ and variance $\Sigma_{i=0}^{n-1} c^{2i}\sigma_{n-i}^2$.

30. An Experimental Method for Obtaining Random Digits and Permutations. J. E. WALSH, U. S. Naval Ordnance Test Station, China Lake.

This paper presents an easily applied method for obtaining small numbers of random binary digits and random permutations. The procedure consists in flipping ordinary minted coins and combining the results of the flips in an appropriate manner. Digits and permutations obtained according to the method of this paper can be considered sufficiently random for any practical application. It appears likely that these digits and permutations are much more nearly random than most of those now available in printed tables. Moreover, any possibility of bias from misuse of tables is avoided. The method presented is particularly suitable for use with respect to experimental designs. Only a few random permutations are ordinarily required for a given experimental design.

31. Distribution of Canonical Partial Correlations. S. N. ROY, University of North Carolina.

By certain general arguments the distribution of canonical partial correlations in random samples of size $n + 1$ from a $(p + q + r)$ variate normal population ($p \leq q, p + q + r \leq n$) can be shown to be of the same form as that of canonical correlations in random samples of size $n + 1 - r$, and involves as parameters (on the non-null hypotheses) the p roots (all lying between 0 and 1) of the equation in θ.

$$|\, \theta(\Sigma_{11} - \Sigma_{12}\Sigma_{33}^{-1}\Sigma'_{12}) - (\Sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma'_{23})(\Sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma'_{23})^{-1}(\Sigma'_{12} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma'_{13}) \,| = 0,$$

where the population co-variance matrix Σ (supposed to be p.d.) is partitioned in the same manner as the sample covariance matrix S of Abstract 17.

In the abstract "On judging all contrasts in the analysis of variance" by Henry Scheffé (*Annals of Math. Stat.*, Vol. 23 (1952), p. 477) the equation $\Sigma_1^k c_i = 0$ was printed incorrectly (due to a compositor's error) as $\Sigma_1^k c_i \theta_i = 0$ on line 5.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Fred C. Andrews has been appointed a Research Associate in the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

Edward W. Barankin, Assistant Professor at the Statistical Laboratory, University of California, Berkeley, has been promoted to Associate Professor. For the academic year 1952/53, Dr. Barankin will be on leave, working at the Institute for Numerical Analysis, Los Angeles.

Z. W. Birnbaum, who has been on leave from the University of Washington for the academic year 1951–1952 and had a visiting professorship in the Department of Statistics at Stanford University, has returned to resume his duties at the University of Washington.

Dr. K. A. Bush, formerly Associate Professor of Mathematics at State University of New York, Champlain College, has accepted an appointment as Assistant Professor of Mathematical Statistics at the University of Illinois.

Charles W. Dunnett, who has been granted a leave as Biometrician for the Food and Drug Laboratories, Ottawa, Canada, has joined the Department of Mathematics, Cornell University for the academic year 1952–3.

Franklin Graybill has recently received his Ph.D. degree from Iowa State College and accepted a position at Oklahoma Agriculture and Mechanical College as Assistant Professor of Mathematics and Associated Statistician to the Agriculture Experiment Station and Research Foundation.

John Gurland, formerly with the Cowles Commission and the Committee on Statistics at the University of Chicago, has joined the staff of the Statistical Laboratory at Iowa State College in Ames.

Wayne W. Gutzman is on leave of absence from the University of South Dakota since being recalled for active duty in the United States Navy and is a Lieutenant Commander and acting Officer-in-Charge of the Computation Ballistics Department of the Naval Proving Ground, Dahlgren, Virginia.

Harman L. Harter, for the past three years an Assistant Professor of Mathematics at Michigan State College, has accepted a position as a Mathematical Statistician at Wright Air Development Center, Wright-Patterson Air Force Base, Dayton, Ohio.

Harry M. Hughes, Instructor at the Statistical Laboratory, University of California, Berkeley, has been promoted to Assistant Professor.

T. J. Jaramillo, formerly Actuary of the Philippine-American Life Insurance Company of Manila, has been appointed as Senior Scientist of the Division of Engineering Mechanics Research of the Armour Research Foundation of the Illinois Institute of Technology, Chicago.

Robert H. Matthias has accepted a position in the Research Division, Electrochemicals Department of E. I. duPont de Neman Co., Niagara Falls.

Lincoln E. Moses has accepted a joint appointment at Stanford University, California, as Assistant Professor in the Department of Statistics and Assistant Professor of Public Health and Preventive Medicine in the Medical School.

Bruce D. Mudgett is now Emeritus Professor of Economics of the University of Minnesota and is residing at Thetford, Vermont.

T. Ellison Neal, formerly statistician for the Textile Division of the U. S. Rubber Company, has joined the staff of the Research Department of the newly formed Chemstrand Corporation at Decatur, Alabama.

B. E. Phillips, since the termination of the research phase of the Parsons-Aerojet Company work at Air Force Missile Test Center, Cocoa, Florida, has been transferred to the Facilities Operation Division of the Ralph M. Parsons Company, Frederick, Maryland.

John Schmid, Jr., resigned from the Board of Examiners of Michigan State College to accept a position as research psychologist with the Research Services Division of the Human Resources Research Center at Lackland Air Force Base, San Antonio, Texas.

Benjamin J. Tepping during the academic year 1952–53 will be on leave as statistician in the Bureau of the Census and will be Visiting Lecturer in the Department of Mathematics and Research Associate in the Survey Research Center, University of Michigan.

Dr. Milton E. Terry has resigned his position as Associate Professor of Statistics at the Virginia Polytechnic Institute and has accepted an appointment as Statistician at the Bell Telephone Laboratory, Murry Hill, New Jersey.

Chia Kuei Tsao has accepted a position as instructor in the Department of Mathematics, Wayne University.

Dr. John E. Walsh, formerly a Consultant with the Census Bureau, is now with the Central Evaluation Group—Code 0110, U. S. Naval Ordnance Test Station, Inyokern, China Lake, California.

Lowell A. Woodbury will be employed for the next two years by the Atomic Bomb Casualty Commission at Hiroshima, Japan, as Chief Biostatistician.

Sidney B. Clark

Sidney B. Clark, member of the Institute for six years, died of a heart attack in Washington, D. C., at the age of 33. He received his B. A. degree at George Washington University. At the time of his death Mr. Clark was a statistician in the National Production Authority. He had been a statistical consultant in the Bureau of Agricultural Economics of the Department of Agriculture and had been employed earlier in other government agencies.

The Educational Testing Service is offering for 1953–54 its sixth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $2,500 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Competence in mathematics and psychology is a prerequisite for obtaining these fellowships. The closing date for completing applications is January 16, 1953. Information and application blanks may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

Three $4000 post-doctoral fellowships in statistics are offered for 1953–54 by the University of Chicago. The purpose of these fellowships, which are open to holders of the doctor's degree or its equivalent in research accomplishment, is to acquaint established research workers in the biological, physical, and social sciences with the crucial role of modern statistical analysis in the planning of experiments and other investigative programs and in the analysis of empirical data. The development of the field of statistics has been so rapid that most current research falls far short of attainable standards, and these fellowships (which represent the third year of a five-year program supported by The Rockefeller Foundation) are intended to help reduce the lag by giving statistical training to scientists whose primary interests are in substantive fields rather than in statistics itself. The closing date for applications is February 1, 1953; instructions for applying may be obtained from the Committee on Statistics, University of Chicago, Chicago 37, Illinois.

New Members

The following persons have been elected to membership in the Institute

(May 31, 1952 to August 29, 1952)

Boggs, Arthur B., M.A. (University of Michigan), Graduate Student, Department of Mathematics, Michigan State College, *136 Albert Avenue, East Lansing, Michigan.*

Borden, Nathan B., M.S. (University of Michigan), Graduate Student, University of Michigan, *1925 Winston Avenue, Louisville 5, Kentucky.*

Broderick, Timonthy S., M.A. (Trinity College, Dublin), Professor of Mathematics, Trinity College, Dublin, *St. Kevin's, Sorrento Road, Dalkey, Co. Dublin, Ireland.*

Eckler, A. Ross, M.A. (Princeton), Graduate Student and Assistant, Mathematics Department, Princeton University, *225-C King Street, Princeton, New Jersey.*

Herd, G. Ronald, M.A. (Kansas University), Chief Statistician and Supervisor of Statistics Department, Aeronautical Radio, Inc., Military Tube Project, *469 Hampton Court, Tyler Gardens, Falls Church, Virginia.*

Hossain, Khondkar Manwar, M.A. (Dacca University, Pakistan), Research Student in Statistics, London School of Economics and Political Science, *Pigeon Hole "H", Research Common Room, London School of Economics and Political Science, Houghton Street, Aldwych, W.C. 2, England.*

Khamis, Salem H., Ph.D. (University of London), Statistician, United Nations, New York, *Statistical Office, United Nations, P.O.B. 20, Grand Central Post Office, New York, N. Y.*

LeCam, Lucien M., Ph.D. (University of California), Instructor, University of California, Department of Mathematics, *Statistical Laboratory, University of California, Berkeley 4, California.*

Morris, Leo E., M.S. (University of Washington, Seattle), Analytical Statistician, Quality Evaluation Laboratory, U. S. Naval Ammunition Depot, Bangor, Washington, *2013 Parkside Drive, Bremerton, Washington.*

Parzen, Emanuel, M.A. (University of California), Research Assistant in Mathematics, University of California, *Department of Mathematics, University of California, Berkeley 4, California.*

Ransom, William E., B.S. (University of Illinois), Statistician, Evaluation and Quality Control Branch, Ordnance Ammunition Center, Joliet, Illinois, *1606 New Lenox Road, Joliet, Illinois.*

Rowan, Michael B., A.B. (George Washington University), Graduate Student, *358 N. Washington Street, Falls Church, Virginia.*

Rutledge, Robert W., B.S. (Sydney), Head Chemist, Pyrmont Distillery, *c/o The Colonial Sugar Refining Company, Ltd., Pyrmont Distillery, Pyrmont, New South Wales.*

Sarhan, Ahmed E. E., B.S. (Fonad I University, Cairo), Statistician, Medical Research Laboratories, Cairo, *1 Clydesdale Rd., Hoylake, Cheshire, England.*

Storer, Robert L., B.S. (University of Nebraska), Statistician, Chief, Evaluation and Quality Control Branch, Ordnance Ammunition Center, *718 John St., Joliet, Illinois.*

Wormleighton, Ralph, M.A. (Toronto), Defence Research Service Officer, *C.A.R.D.E., P.O. Box 1427, Quebec, P.Q., Canada.*

Zindler, Hans-Joachim, Diplom-Mathematiker (University of Göttingen), Scientific Assistant, State Office for Automotive Vehicles, *(24 b) Flensburg, Jurgensgaarderstr. 57 British Zone, Germany.*

REPORT OF THE EAST LANSING MEETING OF THE INSTITUTE

The fifty-third meeting and the fourteenth summer meeting of the Institute of Mathematical Statistics was held in East Lansing, Michigan, at Michigan State College on September 2–5, 1952. The meeting was held in conjunction with meetings of the Mathematical Association of America, the American Mathematical Society, the Econometric Society, and the Pi Mu Epsilon fraternity. Two sessions were co-sponsored by the Econometric Society, and three sessions were co-sponsored by the American Meteorological Society and by the American Geophysical Union. The following 119 members of the Institute attended:

O. P. Aggarwal, G. E. Albert, C. B. Allendoerfer, R. L. Anderson, T. W. Anderson, K. J. Arnold, J. L. Bagg, W. D. Baten, R. E. Bechhofer, M. H. Belz, T. A. Bickerstaff, Allan Birnbaum, David Blackwell, C. R. Blyth, R. C. Bose, G. W. Brier, J. C. Brixey, G. W. Brown, J. H. Bushey, Enrique Cansado, Osmer Carpenter, R. H. Cole, T. F. Cope, A. H. Copeland, E. L. Cox, J. W. Coy, C. C. Craig, D. A. Darling, W. J. Dixon, J. L. Doob, P. S. Dwyer, Benjamin Epstein, H. P. Evans, Evelyn Fix, E. A. Fosler, J. S. Frame, D. A. S. Fraser, Bernard Friedman, J. E. Garrett, H. M. Gehman, M. A. Girshick, R. K. Haddad, S. T. Hadden, P. C. Hammer, M. H. Hansen, T. E. Harris, M. H. Henry, G. R. Herd, Clifford Hildreth, J. L. Hodges, Jr., Wassily Hoeffding, R. G. Hoffman, R. V. Hogg, Jr., W. C. Hood, Harold Hotelling, H. S. Houthakker, C. C. Hurd, P. K. Ito, W. W. Jacobs, T. A. Jeeves, W. H. Jones, Leo Katz, S. H. Khamis, W. M. Kincaid, L. A. Knowler, T. C. Koopmans, William Kruskal, H. G. Landau, L. M. LeCam, E. L. Lehmann, G. J. Lieberman, G. F. Lunger, H. B. Mann, F. J. Massey, Jr., J. W. Mauchly, K. O. May, D. M. Mesner, Robert Mirsky, F. C. Mosteller, C. J. Nesbitt, Jerzy Neyman, M. L. Norden, E. G. Olds, Ingram Olkin, Richard Otter, James Pachares, Emanuel Parzen, G. B. Price, Mina Rees, P. R. Rider, F. D. Rigby, D. D. Rippe, Murray Rosenblatt, S. N. Roy, Herman Rubin, David Rubinstein, R. I. Savage, Henry Scheffé, E. D. Schell, Elizabeth Scott, Esther Seiden, L. D. Simmons, W. B. Simpson, Rosedith Sitgreaves, Arthur Stein, C. M. Stein, L. M. Steinberg, Zenon Szatrowski, W. F. Taylor, H. C. S. Thom, F. H. Tingey, Leo Tornqvist, C. K. Tsao, A. W. Tucker, J. W. Tukey, D. F. Votaw, Jr., Allen Wallis, J. Ernest Wilkins, Jr., B. J. Winer, M. A. Woodbury.

The meeting opened on Tuesday, September 2, at 10:00 A.M. with a session on *Stochastic Phenomena in Medicine*. The chairman was Professor Frederick Mosteller, Harvard University, and the following papers were given:

1. *Some Stochastic Procedures Applied to Data on Tuberculin and Histoplasmin Skin Tests.* William F. Taylor, School of Aviation Medicine, Randolph Field.
2. *Further Contributions to the Theory of Contagion.* Grace E. Bates, Mount Holyoke College and University of California. (Read by Evelyn Fix, University of California.)

3. *An Application of Markov Processes to Problems of Incidence and Epidemiology of Mental Disease.* Andrew Marshall and Herbert Goldhamer, Rand Corporation. (Presented, in outline, by T. E. Harris, Rand Corporation.)

At 2:00 P.M. on the same day a session on *Recent Developments in Measurement in the Social Sciences* was held with Professor Leo Katz, Michigan State College, as chairman. The following papers were given:

1. *Testing Organization Theories.* M. M. Flood, Rand Corporation.
2. *Asymptotic Distributions of Estimates for Latent Structure Analysis.* T. W. Anderson, Columbia University.
3. *Problem of Social Choice and Individual Values.* Leo A. Goodman, University of Chicago.

Discussion was by Professor John W. Tukey, Princeton University, and Professor Max A. Woodbury, University of Pennsylvania.

A session of contributed papers was held at 10:15 A.M. on Wednesday, September 3, with Professor Leo A. Goodman, University of Chicago, as chairman. The following papers were delivered:

1. *An Extension of Massey's Distribution of the Maximum Deviation between Two Sample Cumulative Step Functions. Preliminary Report.* Chia Kuei Tsao, Wayne University.
2. *Polynomial Correlation Coefficients.* W. D. Baten and J. S. Frame, Michigan State College.
3. *Truncated Poisson Distributions.* Paul R. Rider, Wright-Patterson Air Force Base and Washington University.
4. *Frequency Distributions for Functions of Rectangularly Distributed Random Variables.* Stuart T. Hadden, Socony-Vacuum Laboratories, Paulsboro, New Jersey.
5. *On Truncated Rules of Action. Preliminary Report.* Benjamin Epstein, Wayne University.
6. *The Distribution of the Difference of Two Independent Chi-Squares.* James Pachares, University of North Carolina.
7. *Partially Balanced Designs with Two Plots Per Block.* R. C. Bose, University of North Carolina, and K. R. Nair, University of North Carolina and Forest Research Institute, Dehradun, India.
8. *Minimax Sampling and Estimation in Finite Populations.* Om Prakash Aggarwal, Stanford University.
9. *Some Two-Sample Tests on the Exponential Distribution. Preliminary Report.* (By title.) Benjamin Epstein and Chia Kuei Tsao, Wayne University.
10. *Efficiency of Estimators of the Mean of an Exponential Distribution Based Only on the rth Smallest Observation in an Ordered Sample.* (By title.) Benjamin Epstein, Wayne University.
11. *On the Theory of Systematic Sampling. III.* (By title.) William G. Madow, University of Illinois.
12. *The Power of Some Service Tests.* (By title.) Leo A. Goodman, University of Chicago.

A Special Invited Paper was given by Professor Harold Hotelling, University of North Carolina, at 2:00 P.M. on September 3. The title of Professor Hotelling's address was *Distribution of Quadratic Forms.* Professor P. S. Dwyer, University of Michigan, presided.

A session on *Recent Developments in Estimation and Hypothesis Testing in the*

Nonparametric Case was held at 3:00 P.M. on September 3 with Professor W. J. Dixon, University of Oregon, as chairman. The following papers were presented:

1. *The Power of Nonparametric Tests.* Erich L. Lehmann, University of California, Berkeley.
2. *The Power of Certain Nonparametric Tests.* Wassily Hoeffding, University of North Carolina.
3. *Nonparametric Theory: Confidence Regions and Tests for Location and Scale Parameters.* D. A. S. Fraser, University of Toronto.

Discussion was by Professor J. W. Tukey, Princeton University, and Mr. I. R. Savage, National Bureau of Standards.

The three sessions of Thursday, September 4, dealt with problems of *Cloud Seeding* and were co-sponsored by the American Meteorological Society and the American Geophysical Union. The morning session, held at 9:30 A.M., was entitled *Cloud Seeding: The Problem* and had as its chairman Professor Harold Hotelling, University of North Carolina. The following papers were given:

1. *Cloud Seeding; a Problem of National Importance.* R. R. Reynolds, Division of Water Resources, Sacramento.
2. *Physical Basis of Cloud Seeding.* H. G. Houghton, Department of Meteorology, Massachusetts Institute of Technology.
3. *The Physics of Cloud Seeding and the Results of Laboratory Experiments.* Vincent J. Schaefer, General Electric Company, Schenectady.

The afternoon session began at 2:00 P.M. and was a *Review of Already Published Evaluations of Cloud Seeding Experiments.* The chairman was Dr. Howard T. Orville, Bendix Aviation Corporation, and the following papers were given:

1. *Evaluation of the Bishop Creek, California, Cloud Seeding Tests.* Ferguson Hall, Scientific Services Division, Weather Bureau.
2. *Methods of Evaluating the Effects of Periodic Silver Iodide Seeding.* Irving Langmuir, General Electric Company, Schenectady.
3. *Progress of Cloud Seeding Analysis in Oregon.* Robert T. Beaumont, Bureau of Soil Conservation, Medford, Oregon.
4. *Some Pitfalls Encountered in Certain Current Methods of Evaluation.* T. A. Jeeves, L. LeCam, E. L. Scott, University of California, Berkeley.
5. *An Approach to the Evaluation of Results of Rainmaking, A Progress Report.* C. E. Buell, University of New Mexico.

In the evening, at 7:00 P.M., a session was held with the title *Proposed Statistical Methodology and Round Table Discussion.* The chairman was Professor Henry Scheffé, Columbia University, and the following papers were given:

1. *On Proposed Statistical Methodology.* J. Neyman, University of California, Berkeley.
2. *Methods of Evaluating Cloud Seeding Operations.* Herbert C. S. Thom, Weather Bureau and Cornell University.
3. *Statistical Problems Encountered in the Analysis of Cloud Seeding Experiments.* Glenn W. Brier, Statistical Section, Weather Bureau.

Each of the three Cloud Seeding sessions was followed by a discussion.

On Friday, September 5, at 10:00 A.M. a second session of contributed papers was held with Professor C. C. Craig, University of Michigan, presiding. The following papers were presented:

1. *A Minimal Essentially Complete Class of Tests of a Simple Hypothesis Specifying the Mean of a Unit Rectangular Distribution.* Allan Birnbaum, Columbia University.
2. *Application of Random Walk Theory to a General Class of Sequential Decision Problems. Preliminary Report.* G. E. Albert, University of Tennessee.
3. *Nonparametric Comparisons of Populations When Data Are Collected in Homogeneous Groups.* Frank J. Massey, Jr., University of Oregon.
4. *On the Reduced Moment Problem.* Salem H. Khamis, Statistical Office, United Nations.
5. *Canonical Partial Correlations.* S. N. Roy and J. Whittlesey, University of North Carolina.
6. *A Useful Transformation in the Case of Canonical Partial Correlations.* S. N. Roy, University of North Carolina.
7. *Uniform Convergence of Distribution Functions.* Emanuel Parzen, University of California, Berkeley.
8. *Statistical Aspects of a Linear Programming Problem.* D. F. Votaw, Jr., Yale University.
9. *Maximum Likelihood Estimators and A Posteriori Distributions.* (By title.) J. Wolfowitz, Cornell University.
10. *Estimates and Asymptotic Distributions of Certain Statistics in Information Theory. Preliminary Report.* (By title.) John P. Hoyt, U. S. Naval Academy. Introduced by S. Kullback.
11. *On Testing One Simple Hypothesis against Another.* (By title.) Lionel Weiss, University of Virginia.
12. *Extreme Value Theory for m-dependent Stationary Sequences of Continuous Random Variables.* (By title.) Geof. Watson, University of Melbourne.
13. *Sequential Tests and Estimates for Comparing Poisson Populations.* (By title.) Allan Birnbaum, Columbia University.
14. *Sequential Decision Problems in the Stationary Case.* (By title.) J. Kiefer, Cornell University.
15. *Random Functions Satisfying Certain Linear Relations. II.* (By title.) Sudhish G. Ghurye, University of North Carolina.
16. *Optimal Designs for Estimating Parameters. Preliminary Report.* (By title.) Herman Chernoff, Stanford University.
17. *The Distribution of the nth Variate in Certain Chains of Serially Dependent Populations.* (By title.) L. V. Toralballa, Marquette University. Introduced by Joseph Talacko.
18. *An Experimental Method for Obtaining Random Digits and Permutations.* (By title.) J. E. Walsh, U. S. Naval Ordnance Test Station, China Lake.
19. *Distribution of Canonical Partial Correlations.* (By title.) S. N. Roy, University of North Carolina.

At 2:00 P.M. on Friday, September 5, a session on the *Comparison of Experiments* was held with the co-sponsorship of the Econometric Society. The chairman was Professor M. A. Girshick, Stanford University, and the following papers were presented:

1. *Equivalent Methods of Comparison.* David H. Blackwell, Howard University.
2. *Approximate Comparison of Experiments.* Charles Stein, University of Chicago.

At 3:00 P.M. on Friday, September 5, the final session of the meeting was held. This session was one of invited papers, and its chairman was Professor J.

Neyman, University of California, Berkeley. The invited addresses and discussions were the following:

1. *Estimates of Bounded Relative Error in Particle Counting.* (Based on joint work of M. A. Girshick, H. Rubin, and Rosedith Sitgreaves.) Rosedith Sitgreaves, Stanford University.
2. *Multiple Comparison Procedures in the Analysis of Variance.* Henry Scheffé, Columbia University.
   Discussion: Robert Bechhofer, Cornell University, and J. W. Tukey, Princeton University.
3. *Topics in Stationary Time Series.* (Based on joint work with Murray Rosenblatt.) Ulf Grenander, University of Stockholm and University of Chicago.
   Discussion: Murray Rosenblatt, University of Chicago.

The Council met at 8:00 P.M. on Tuesday, September 2, and a Business Meeting was held at 9:00 A.M. on Wednesday, September 3. At both of these meetings Professor M. A. Girshick presided. A banquet was held on the evening of September 3, and an I.M.S. party on the evening of September 4.

WILLIAM KRUSKAL
Associate Secretary

PUBLICATIONS RECEIVED

Anuario Estadistico de España, (Instituto Nacional de Estadistica), Presidencia del Gobierno, Madrid, 1951, xlv + 1258 pp.

Anuario Estadistico de España (Edicion Manual), (Instituto Nacional de Estadistica), Presidencia del Gobierno, Madrid, 1952, lvii + 896 pp.

INSTITUTIONAL MEMBERS

The following are Institutional Members of the Institute for the year 1952:

INTERNATIONAL BUSINESS MACHINES CORPORATION, New York
UNIVERSITY OF NORTH CAROLINA, INSTITUTE OF STATISTICS, Chapel Hill, North Carolina
PRINCETON UNIVERSITY, DEPARTMENT OF MATHEMATICS, SECTION OF MATHEMATICAL STATISTICS, Princeton, New Jersey
PURDUE UNIVERSITY LIBRARIES, Lafayette, Indiana
RAYTHEON MANUFACTURING COMPANY, 148 California Street, Newton 58, Massachusetts
UNIVERSITY OF CALIFORNIA, STATISTICAL LABORATORY, Berkeley, California
UNIVERSITY OF ILLINOIS, PERIODICAL DIVISION, Urbana, Illinois

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]